Leonard Bolc
Ryszard Tadeusiewicz
Leszek J. Chmielewski
Konrad Wojciechowski (Eds.)

LNCS 6374

Computer Vision and Graphics

International Conference, ICCVG 2010
Warsaw, Poland, September 2010
Proceedings, Part I

1 Part I

Springer

Lecture Notes in Computer Science 6374

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Leonard Bolc Ryszard Tadeusiewicz
Leszek J. Chmielewski
Konrad Wojciechowski (Eds.)

Computer Vision and Graphics

International Conference, ICCVG 2010
Warsaw, Poland, September 20-22, 2010
Proceedings, Part I

Springer

Volume Editors

Leonard Bolc
Konrad Wojciechowski
Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008 Warsaw, Poland,
E-mail: {pjwstk, konradw}@pjwstk.edu.pl

Ryszard Tadeusiewicz
AGH University of Science and Technology, Institute of Automatics
Al. Mickiewicza 30, 30-059, Kraków, Poland
E-mail: rtad@agh.edu.pl

Leszek J. Chmielewski
Warsaw University of Life Sciences
Faculty of Applied Informatics and Mathematics
Department of Informatics
Nowoursynowska 159, 02-776 Warsaw, Poland
E-mail: leszek_chmielewski@sggw.pl

Konrad Wojciechowski
The Silesian University of Technology, Institute of Computer Science
Akademicka 16, 44-100 Gliwice, Poland
E-mail: konrad.wojciechowski@polsl.pl

Library of Congress Control Number: 2010934223

CR Subject Classification (1998): I.4, I.2.10, I.5, I.3.5, I.2, I.3

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

ISSN 0302-9743
ISBN-10 3-642-15909-5 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-15909-1 Springer Berlin Heidelberg New York

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

The International Conference on Computer Vision and Graphics, ICCVG, organized since 2002, is the continuation of the International Conferences on Computer Graphics and Image Processing, GKPO, held in Poland every second year from 1990 to 2000. The founder and organizer of these conferences was Prof. Wojciech Mokrzycki. The main objective of ICCVG is to provide a forum for the exchange of ideas between researchers in the closely-related domains of computer vision and computer graphics.

ICCVG 2010 gathered about 100 authors. The proceedings contain 95 papers, each accepted on the grounds of two independent reviews. During the conference two special sessions were organized: *Advances in Pattern Recognition, Machine Vision and Image Understanding* and *Human Motion Analysis and Synthesis.*

The content of the issue has been divided into three parts. The first and second parts are related to the two special sessions mentioned above, containing 11 chapters each. The third part, named like the whole book, Computer Vision and Graphics, collects all the remaining chapters.

ICCVG 2010 was organized by the Association for Image Processing, Poland (Towarzystwo Przetwarzania Obrazów – TPO), the Polish-Japanese Institute of Information Technology (PJWSTK), and the Faculty of Applied Informatics and Mathematics, Warsaw University of Life Sciences (WZIM SGGW).

The Association for Image Processing integrates the Polish community working on the theory and applications of computer vision and graphics. It was formed between 1989 and 1991.

The Polish-Japanese Institute of Information Technology, founded in 1994 by the Computer Techniques Development Foundation under the agreement of the Polish and Japanese governments, is one of the leading, non-state (private) Polish universities. We are highly grateful for the fact that the institute has been hosting and supporting the Conference.

The Faculty of Applied Informatics and Mathematics, established in 2008 at the Warsaw University of Life Sciences, offers two programs of studies: Informatics, and Informatics and Econometrics. Its main advantage is its focus on merging technical education with applied sciences, including application of computer sciences to the management and analysis of the agricultural industry.

We would like to thank all the members of the Scientific Committee, as well as the additional reviewers, for their help in ensuring the high quality of the papers. We would also like to thank Grażyna Domańska-Żurek for her excellent work on technically editing the proceedings, and Bernadeta Bonio and Paweł Wieman for their engagement in the conference organization and administration.

September 2010

Leonard Bolc
Ryszard Tadeusiewicz
Leszek J Chmielewski
Konrad Wojciechowski

Organization

- Association for Image Processing (TPO)
- Polish-Japanese Institute of Information Technology (PJWSTK)
- Faculty of Applied Informatics and Mathematics,
  Warsaw University of Life Sciences (WZIM SGGW)

Conference General Chairs

J.L. Kulikowski (Poland)
L.J. Chmielewski (Poland)
K. Wojciechowski (Poland)

Conference Committee Members

A. Bagdanov (Spain)
I. Bajla (Austria)
E. Bengtsson (Sweden)
M. Berndt-Schreiber (Poland)
P. Bhattacharya (USA)
G. Borgefors (Sweden)
A. Borkowski (Poland)
D. Chetverikov (Hungary)
L. Chmielewski (Poland)
L. Czuni (Hungary)
S. Dellepiane (Italy)
M. Domański (Poland)
M. Flasiński (Poland)
B. Gabryś (UK)
A. Gagalowicz (France)
G. Giacinto (Italy)
D. Gillies (UK)
E. Hancock (UK)
L. Heutte (France)
V. Hlavac (Czech Republic)
M. Iwanowski (Poland)
A. Jóźwik (Poland)
H. Kälviäinen (Finland)
A. Kasiński (Poland)
J. Kacprzyk (Poland)
W. Kasprzak (Poland)
R. Kasturi (USA)
N. Kiryati (Israel)
R. Klette (New Zealand)
J. Korbicz (Poland)
R. Kozera (Australia)
K. Krawiec (Poland)
H.J. Kreowski (Germany)
A. Krzyżak (Canada)
M. Kurzyński (Poland)
B. Kwolek (Poland)
Y.B. Kwon (South Korea)
A. Materka (Poland)
N. Mavridis (United Arab Emirates)
W. Mokrzycki (Poland)
H. Niemann (Germany)
M. Nieniewski (Poland)
S. Nikiel (Poland)
L. Noakes (Australia)
A. Nowakowski (Poland)
M. Ogiela (Poland)
M. Orkisz (France)
R.M. Palenichka (Canada)
H. Palus (Poland)
W. Pamula (Poland)
F.J. Perales López (Spain)
J. Piecha (Poland)
E. Piętka (Poland)
P. Porwik (Poland)
A. Przelaskowski (Poland)
J. Punys (Lithuania)
J. Roerdink (The Netherlands)
P. Rokita (Poland)
L. Rutkowski (Poland)
K. Saeed (Poland)
G. Sanniti di Baja (Italy)
R. Šára (Czech Republic)
G. Schaefer (UK)
B. Scotney (Ireland)
V. Skala (Czech Republic)
W. Skarbek (Poland)
A.N. Skodras (Greece)
G. Sommer (Germany)
C. Stępień (Poland)
A. Śluzek (Singapore)
R. Tadeusiewicz (Poland)
J.M.R.S. Tavares (Portugal)
K. Tombre (France)
E. Derya Übeyli (Turkey)
V. Valev (Bulgaria and Saudi Arabia)
W. Malina (Poland)
M. Woźniak (Poland)
Z. Wróbel (Poland)
J. Zabrodzki (Poland)
W. Zamojski (Poland)
M. Zaremba (Canada)

Table of Contents – Part I

Advances in Pattern Recognition, Machine Vision and Image Understanding

Human Motion Analysis and Synthesis

Computer Vision and Graphics

Table of Contents – Part II

Computer Vision and Graphics

Visual Codebooks Survey for Video On-Line Processing

Vítězslav Beran and Pavel Zemčík

Department of Computer Graphics and Multimedia
Faculty of Information Technology, Brno University of Technology
Božetěchova 2, 612 66 Brno, CZ
{beranv,zemcik}@fit.vutbr.cz

Abstract. This paper explores techniques in the pipeline of image description based on visual codebooks suitable for video on-line processing. The pipeline components are *(i)* extraction and description of local image features, *(ii)* translation of each high-dimensional feature descriptor to several most appropriate visual words selected from the discrete codebook and *(iii)* combination of visual words into bag-of-words using hard or soft assignment weighting scheme. For each component, several state-of-the-art techniques are analyzed and discussed and their usability for video on-line processing is addressed. The experiments are evaluated on the standard Kentucky and Oxford building datasets using image retrieval framework. The results show the impact loosing the pipeline precision in the price of improving the time cost which is crucial for real-time video processing.

1 Introduction

The main interest of this work is the exploration of image retrieval techniques based on visual codebooks for their utilization for video on-line processing. The image retrieval frameworks based on visual codebooks are mainly composed from several components. The local visual features are extracted from the images and described using high-dimensional descriptors. Having pre-trained discrete visual codebook, the high-dimensional descriptors are translated to several most appropriate *visual words*. The images are then represented as the distribution of these visual words that is called bag-of-words.

Described approach is inspired by text retrieval systems. One of the first work introducing visual codebooks was *Video Google* [1]. The work utilized widely used SIFT transformation [2] as local features. The visual vocabulary training was based on naive k-means algorithm. When introduced, the visual codebooks were used in image and object retrieval applications. Similarly to text retrieval, the database of images is indexed by inverted file approach. Search for a query image in database results in immediate returning of a ranked list of documents (key frames, shots, etc.) in the manner of search in text documents.

Later works utilized more types of local image features such as corner-like detector with full affine adaptation or detection of stable regions. When the

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 1–10, 2010.

local features are finally described, the performance of the visual vocabulary approach correlates to performance of particular image feature extraction technique. Needs for large, more discriminative vocabularies lead the research to find faster clustering methods. Two significant methods were developed: hierarchical k-means [3] and approximated k-means [4]. The developed methods allow creation of vocabularies with size about 1M of visual words with reasonable time and computational cost.

The later research experimenting with different translation schemas when translating the local image features into bag-of-words showed that the discriminative power of the vocabulary could be improved not only by extending the vocabulary size. Instead of the standard approach, where one local feature is translated in just one visual word, a single image feature could be assigned to several visual words. The approach is known as *soft-weighting*.

The existing works mainly explore the solutions for their best *precision* performance. The work [5] compares several detectors of local image features and evaluate their performance on visual codebooks with different sizes. The explored detectors are time expensive and the codebook sizes are only up to 10k. Other solution [6] also offers the results of comparison of visual codebooks with different size (up to 1M), but also uses the time expensive image feature detector solution. The work [7] is focused on time effectiveness. Standard SIFT and SURF feature extractors are optimized for dense sampling and descriptors dimensionality is reduced by PCA.

This paper analyzes the state-of-the-art techniques used in each pipeline component in the utilization in video on-line processing point of view. Next Section contains an overview of local image feature extraction and description methods explaining their crucial attributes for real-time systems. The process of building the visual codebook is described in Section 3 together with codebook searching methods. The possibilities of visual words weighting when bag-of-words are constructed are discussed in Section 4. Sections 5 and 6 describe the experiments and discuss the results.

2 Image Local Features

Real-time applications such as on-line video synchronization introduce specific demands to the commonly used techniques. The attributes of image local feature extraction methods are stability, repeatability and robustness to several types of transformations or distortions [8]. The characteristic of feature descriptor is its discriminative power. Usually, the more powerful the feature extraction and description methods are, the higher is their time cost. For the real-time applications dealing with consecutive video-frame processing, the methods performance could be decreased at the expense of execution time increase. The computational cost demands also deriving the size of the visual vocabulary. Detected local image features are expected to be invariant to geometric and illumination changes. Different detectors emphasize different aspects of invariance, resulting in keypoints of varying properties and sampled sizes.

Widely used method is SIFT [2]. The method is sensitive to blob-like structure and is invariant to scale and orientation. Besides the local feature detection, SIFT also describes the local regions using histogram of gradients. The SIFT preserves the gradient location information by dividing the region into regular grid of 4x4 subregions. SIFT is often used only as a descriptor for different image local feature detectors. The blob-like structures can be also detected using Hessian matrix [8]. The approach based on Hessian matrix can be effectively approximated by block filters (SURF [9]). The SURF detector is based on effective platform computation of Haar-wavelets on integral images. The authors also introduced new descriptors utilizing the same platform. The approach known as *FAST corners* ([10]) employs machine learning to construct a corner detector that outperforms all know approaches in the speed point of view. The FAST itself is neither invariant to scale nor to shear. When full affine invariance is necessary, the characteristic scale selection and affine adaptation [11] can be applied. Unfortunately, full affine transformation detection significantly slows down the process. Detected corners are described using a gradient distribution of the region around the detected point. Gradient distribution can be described by Histogram of Gradients (*HOG* [12]) or Gradient Location and Orientation Histogram (*GLOH* [13]).

The FAST detector combined with the GLOH or HOG descriptor is not rotational invariant. This paper present some modification of these descriptors to improve their robustness to rotation transformation.

3 Visual Codebooks

The idea of visual vocabulary, firstly used in *Video Google* by [1], brings the techniques from natural language processing and information retrieval area. The document (image) is represented as an unordered collection of words (bag-of-words model). In computer vision, the (visual) words might be obtained from the feature vectors by a quantization process. The objective is to use vector quantization to descriptors to translate them into clusters' labels which represents the visual words.

Visual vocabulary is built during the training stage. A part of the data (training data) is used to divide the descriptor space into clusters. Each cluster is labeled; has its own identification number. The vocabulary is then the list of cluster centers and identifiers. The clustering procedure based on k-mean algorithm contains the search step, when the sample should be assigned to the nearest. The later research introduced several solutions to avoid time consuming naive sequential search. Figure 1 schematically shows the different approaches described in detail below.

When the size of the resulting vocabulary is small ($k < 10^3$), the *naive k-means* algorithm can be used (Fig. 1(a)). The time complexity of the k-means algorithm is $O(kN)$, where N is the number of training feature vectors. Some applications (e.g. for object retrieval [4]) need more discriminative vocabulary. One possible way how to reduce the time complexity is using *Hierarchical k-means*

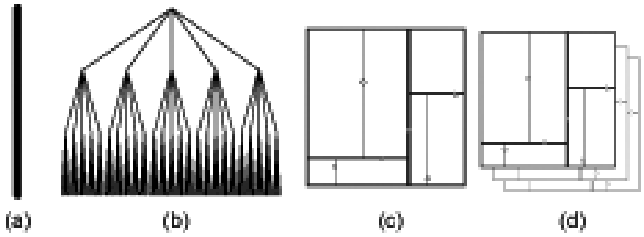


Fig. 1. Clustering strategies - (a) naive sequential, (b) hierarchical, (c) kd-tree and (d) random forest

[3] (Fig. 1(b)). Instead of solving one clustering with a large number of cluster centers, a tree organized hierarchy of smaller clustering problems is solved. This reduces the time complexity to $O(N. \log k)$. The problem with Hierarchical k-means is that it optimizes the problem only locally, per tree branch. Other approach reducing the time complexity is replacing the nearest neighbor search of k-means by *kd*-tree (Fig. 1(c)) or by random forest of *kd*-trees (Fig. 1(d)). This approach is called *Aproximate k-means* [4]. The quantization error after clustering procedure is expressed as a sum of distances of training samples to their nearest cluster as follows:

$$D = \frac{1}{N} \sum_{i=1}^{N} d(p_i, Q[p_i]) \tag{1}$$

where N is the number of training samples Q is the nearest cluster center to the sample p_i and d is the distance function.

Having visual codebook and the dataset, each visual word appears in different amount of images and also different times in each particular image. Some of the visual words are quite rare in contrary to visual words that appears very often. Usually, standard weighting used in text retrieval is employed that is known as 'term frequency - inverse document frequency' - *tf-idf*. The *term frequency* reflects the entropy of a word with respect to each document unlike *inverse document frequency* down-weights words that appear often in the database. The resulting weight is then:

$$\begin{aligned} tf - idf(w) &= tf(w).idf^{log}(w) &(2)\\ &= \frac{|d(w)|}{|d|}. \log(\frac{|D|}{|D(w)|}) &(3) \end{aligned}$$

where d is a document (image signature), $|d|$ is a number of words in d and $|d(w)|$ is the number of occurrences of word w in d, D is a dataset of all documents and $D(w)$ is a set of documents containing the word w.

The *idf* weighting function emphasizes the rare visual words and down-weights the frequent ones. The rare visual words does not necessarily mean that they are best informative. We introduce another weighting function idf^{hat} that is defined as follows:

$$idf^{hat}(w) = \exp\left(-\frac{1}{2}\left(\frac{|D|}{|D(w)} - 1\right)^2 c\right) \tag{4}$$

where c is the steepness of the function experimentally evaluated and suggest $c = 9.0$. The hat-like function down-weights both frequent and rare visual words and emphasizes the common ones.

4 Visual Words Assignment

The *bag-of-words* is a collection of weighted visual words representing the image content. The bag-of-words is also known as *image signature*. This collection can be seen also as a vector of visual word frequencies. It degrades to a set-of-words when the weights represent only the word's presence (binary vector). Otherwise, it is a bag-of-words. The Figure 2 shows the process of describing the image content by an image signature.

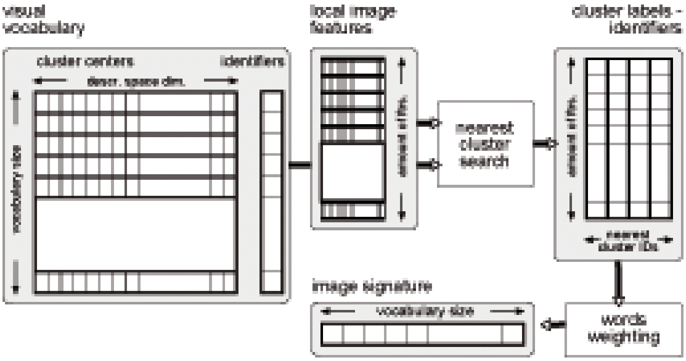


Fig. 2. Image signature extraction procedure

Having visual codebook, for each descriptor of local features from the image, the k visual words (nearest clusters) are found. The weight for each word is computed and used to increase the value of the image signature at the word's ID position. The image signature then can be seen as a histogram of occurred visual words. The works ([5], [6]) reflected the fact that the quantization effect provides a very coarse approximation to the actual distance between two features - zero if assigned to the same visual word and infinite otherwise. Such approach is called *hard assignment*. The *soft-assignment* (soft-weighting) techniques assign a single descriptor to several visual words nearby in the descriptor space. Given the sorted list of k *nearest* visual words, the weighting functions assign different weight to the visual word according to its distance or its rank in the list. The weighting functions can be defined as follows:

$$w_i^{ratio} = \frac{d_0}{d_i} \tag{5}$$

$$w_i^{exp} = \exp(-\frac{d_i^2}{2\sigma^2}) \tag{6}$$

$$w_i^{rank} = \frac{1}{2^{i-1}} \tag{7}$$

where i is the rank and d_i is the distance of the ith descriptor point in the list to its closest visual word. The basic weighting function (Eq 5.) is the ratio between the distance of the descriptor point to the closest visual word and distance of the actual point to its closest visual word. The exponential function (Eq 6.) uses σ so that substantial weight is only assigned to a small number of cells. The authors [6] experimentally evaluated and suggest $k = 3; \sigma^2 = 6.250$. The similar idea is realized by rank function (Eq 7.) replacing the distance by the visual word's rank[5].

During the retrieval stage, documents are ranked by their *similarity*. One of the frequently used *similarity metric* in text retrieval is normalized scalar product (cosine of angle) between the query vector **q** and all document vectors **d** in the database. The cosine similarity can be seen as a method of normalizing document length during comparison and is defined as $sim(\mathbf{q}, \mathbf{d}) = \frac{\mathbf{q}.\mathbf{d}}{|\mathbf{q}||\mathbf{d}|}$ where . is dot product and || is the vector magnitude. The cosine similarity of two image signatures will range from 0 meaning independent images to 1 meaning exactly the same images, since the word weights cannot be negative.

5 Experiments

The image datasets was used for the presented experiments analyzing the characteristics of various parameters of feature extraction and visual vocabulary building. *Kentucky Dataset* was created as a recognition benchmark at Kentucky University [3]. The set consists of 2550 groups of 4 images each, that is 10200 images in total. The objects images are taken from different angles and rotations. The size of the images is approximately 640x480 pixels. More details about the extracted local features is in Table 1.

Table 1. The number of images and features for each dataset

dataset	images	SIFT	SURF	FAST HOG	FAST OGH
Kentucky	10200	13.161.824	6.541.246	2.664.175	5.013.715

The experiments evaluating the clustering process based on k-mean algorithm are designed to measure the clustering error (Eq. 1) in each iteration step. The measurement procedure is repeated several times to analyze the influence of the initialization error.

Next set of experiments are design to analyze the relation between the speed and precision performance. The speed is measured and compared for each pipeline component separately. The precision is measured for the entire retrieval pipeline as a *Mean Average Precision* (mAP). Average Precision represents the area under Precision-Recall curve for a query and can be directly computed from the ranked list of retrieved images as:

$$AP = \frac{1}{m} \sum_{i=1}^{n} \frac{relevant(x_i)}{i} \tag{8}$$

where n is the number of retrieved images, m is the number of relevant images, x_i is the i-th image in the ranked list of retrieved images $X = x_1, \ldots, x_n$ and $relevant(x_i)$ returns the number of relevant images in the first i images, only if the x_i is relevant image itself, and 0 otherwise. This measure gives a number in range (0, 1] where a higher number corresponds to a better performance.

Evaluation of different methods for their usability for video on-line processing is done by experimenting with different sizes of codebook, different clustering strategies, various weighting schemes in bag-of-words and different types of image local feature extractors.

6 Results

Clustering error. The k-mean clustering used to create the visual codebook is an iterative process. The method minimizes the error in each step. The clustering error progress was evaluated in SIFT and SURF descriptor space. The resutls in Fig. 3 shows the considerable clustering improvements up to 4 iterations in average.

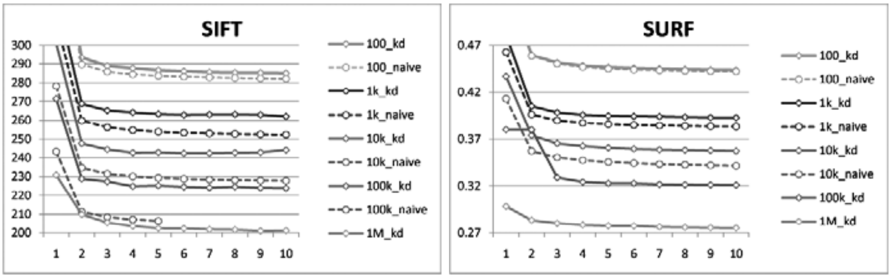


Fig. 3. Clustering error for SIFT and SURF descriptor space

The k-means clustering is known to be sensitive to initialization. The experiments with different clustering initialization result in nearly same clustering error. It approves the hypotheses that the samples in the descriptor space do not gather in significant groups but are rather regularly distributed.

Vocabulary size. The previous works approved the assumption, that bigger codebook gives better results. The designed experiments compares the codebooks with size 100, 1k, 10k, 100k and 1M using two baseline image features SIFT and SURF focusing also on the speed of each run. The results in Fig. 4 show that the naive search strategy outperforms the kd-tree search but the time cost is exponentially higher. Depending on the image data type, the precision of bigger codebooks with naive search reach the precision of smaller ones with kd-tree search. The results are evaluated using idf^{log} and hard-assignment.

size	kd-tree		naive		log(speed)	
	SIFT	SURF	SIFT	SURF	kd-tree	naive
100	0.583	0.525	0.591	0.524	1.21	1.52
1k	0.594	0.550	0.637	0.574	1.21	2.54
10k	0.600	0.551	0.659	0.600	1.47	3.69
100k	0.669	0.569	0.756	0.674	1.76	4.64
1M	0.728	0.614			1.78	

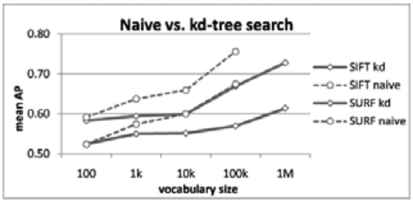


Fig. 4. Performance comparison of the codebooks with different sizes

Codebook weighting. The proposed weighting function for computing inverse document frequency idf^{hat} was compared to standard approach based on logarithm function idf^{log}. The assumption that the enhancing of the rare visual words negatively influence the codebook precision is wrong. The results in Fig. 5 show that idf^{log} outperforms the idf^{hat} function. The performance for codebook with size 1k is caused clearly by soft-assignment approach. The results are evaluated using SURF descriptor and kd-tree search strategy.

size	knn=1		knn=10		
	Log	Hat			
	Ratio0	Ratio0	Ratio0	Exp0	None0
100	0.525	0.472	0.517	0.509	0.498
1k	0.550	0.481	0.572	0.573	0.554
10k	0.551	0.464	0.506	0.510	0.488
100k	0.569	0.510	0.549	0.550	0.518
1M	0.614	0.556	0.607	0.608	0.590

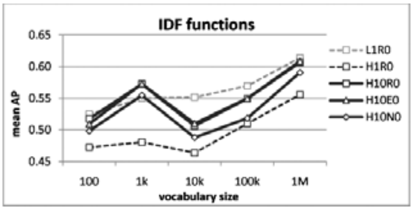


Fig. 5. The number of images and features for each dataset

Soft assignment. The soft-assignment techniques significantly improves the pipeline performance. Several combinations of weighting functions are explored. The abbreviations used in result tables and graphs means - *None* for no weighting, *Exp* for exponential function and *Ratio* for ratio function. These functions are combined with *Rank* function which is marked by number *2* in the name of the experiment run (when *0*, no ranking is used). The number of used visual words for soft-assignment is in the *knn* (k nearest neighbor) column. The experiments in Fig. 6 revealed that no more than 4 closest visual words to the descriptor significantly improved the overall performance.

knn	None0	Exp0	Exp2	Ratio0	Ratio2
1	0.550	0.550	0.550	0.550	0.550
4	0.584	0.606	0.612	0.608	0.609
10	0.588	0.606	0.615	0.609	0.611

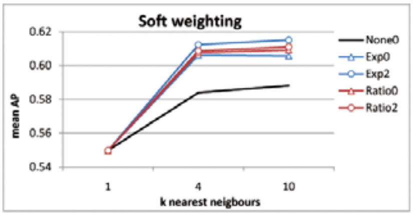


Fig. 6. Soft-assignment functions and their combinations

Image local features. Four image local feature extracts and descriptors were evaluated in experiments - SIFT, SURF, FAST detector combined with HOG descriptor (FHOG) and FAST detector combined with GLOH descriptor (FGLOH). The GLOH descriptor contain modification for descriptor rotation invariance. The results in Fig. 7 show that the SURF has comparable precision as SIFT but is 3 times faster. The speed results are measured as average time of one feature extraction in 10^{-6} seconds.

size	SIFT	SURF	FHOG	FGLOH
1k	0.587	0.609	0.472	0.457
100k	0.649	0.593	0.465	0.445
speed	1.00	3.28	3.65	8.34

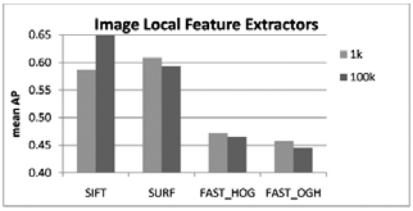


Fig. 7. Speed and performace comparison of different image local feature extraction methods

Balancing the speed and precision performance, the codebook for SURF or FAST+GLOH with 100k visual words using the kd-tree search strategy and soft-assignment yield the usable results for video processing solutions working on-line.

7 Conclusion

The presented work explores the image retrieval pipeline components. The state-of-the-art techniques based on visual codebooks were discussed for their utilization for video on-line processing. The experiments cover the influence of method choice and their parameters configuration for each component. The results show the precisions and speeds of codebooks with various sizes, with different weighting functions computing the visual word entropy, couple of soft-assignment strategies and two different search strategies for clustering method.

References

1. Sivic, J., Zisserman, A.: Video Google: Efficient visual search of videos. In: Ponce, J., Hebert, M., Schmid, C., Zisserman, A. (eds.) Toward Category-Level Object Recognition. LNCS, vol. 4170, pp. 127–144. Springer, Heidelberg (2006)
2. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), 91–110 (2004)
3. Nister, D., Stewenius, H.: Scalable recognition with a vocabulary tree. In: CVPR 2006: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, pp. 2161–2168. IEEE Computer Society, Los Alamitos (2006)
4. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Object retrieval with large vocabularies and fast spatial matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2007)
5. Jiang, Y.G., Ngo, C.W., Yang, J.: Towards optimal bag-of-features for object categorization and semantic video retrieval. In: CIVR 2007: Proceedings of the 6th ACM International Conference on Image and Video Retrieval, pp. 494–501. ACM, New York (2007)
6. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Lost in quantization: Improving particular object retrieval in large scale image databases. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2008)
7. Uijlings, J.R.R., Smeulders, A.W.M., Scha, R.J.H.: Real-time bag of words, approximately. In: CIVR 2009: Proceeding of the ACM International Conference on Image and Video Retrieval, pp. 1–8. ACM, New York (2009)
8. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.V.: A comparison of affine region detectors. Int. J. Comput. Vision 65(1-2), 43–72 (2005)
9. Bay, H., Tuytelaars, T., Gool, L.V.: Surf: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
10. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
11. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. Int. J. Comput. Vision 60(1), 63–86 (2004)
12. Dalal, N., Triggs, B., Schmid, C.: Human detection using oriented histograms of flow and appearance. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 428–441. Springer, Heidelberg (2006)
13. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. 27(10), 1615–1630 (2005)

Application of Shape Description Methodology to Hand Radiographs Interpretation

Marzena Bielecka[1], Andrzej Bielecki[2], Mariusz Korkosz[3], Marek Skomorowski[2],
Wadim Wojciechowski[4], and Bartosz Zieliński[2]

[1] Department of Geoinformatics and Applied Computer Science
Faculty of Geology, Geophysics and Environmental Protection
AGH University of Science and Technology, Mickiewicza 30, 30-059 Cracow, Poland
bielecka@agh.edu.pl

[2] Institute of Computer Science, Jagiellonian University
Łojasiewicza 6, 30-348 Cracow, Poland
uibielec@cyf-kr.edu.pl, skomorowski@ii.uj.edu.pl,
bartosz.zielinski@uj.edu.pl

[3] Division of Rheumatology, Departement of Internal Medicine and Gerontology
Jagiellonian University Hospital, Śniadeckich 10, 31-531 Cracow, Poland
mariuszk@mp.pl

[4] Cracow University Hospital, Department of Radiology
Kopernika 19, 31-531 Cracow, Poland
wadim@mp.pl

Abstract. In this paper, a shape description methodology, introduced by Jakubowski [6] is applied to hand radiographs interpretation, in order to recognize bones borders shapes in the fingers regions. It is shown that the classical approach can be used only for preliminary analysis. Therefore, the improved method, based on fuzzy approach, is considered.

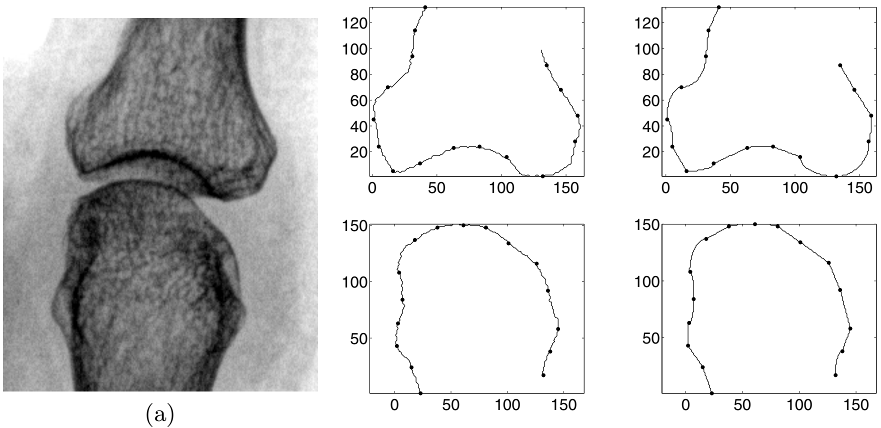
1 Introduction

Hand radiographs are source of important clinical information. One of the most significant radiological features of inflammatory and non-inflammatory diseases is joint surfaces erosion [11]. To give a diagnosis, an X-ray is taken of the patients hand and symmetric metacarpophalangeal joint spaces and interphalangeal joint spaces are analyzed.

The changes in border of metacarpophalangeal joints surfaces observed on hand radiographs are a crucial point in medical diagnosis and support important information for estimation of therapy efficiency. However, they are difficult to detect in an X-ray picture when examined by a human expert, due to the quantity of joints. The possibility of performing such analysis by a computer system is a key point for diagnosis support. Therefore, studies concerning possibilities of implementation such systems are topic of numerous publications [9, 10, 13] (see other references in [4]).

This paper is a continuation of studies described in [1–4, 15, 16]. In the previous papers the preprocessing and joint location algorithms were presented. At

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 11–18, 2010.



(a)

(b c)

(d e)

Fig. 1. Radiograph of the healthy metacarpophalangeal joint (a), borders of its upper (b) and lower (c) surfaces and their respective contours (d, e)

the beginning, the applied approach turned to be effective in about 90% of cases [4], however the algorithm was then improved in [16] and efficiency at 97% was achieved. Based on those locations, the algorithm identifying the borders of the upper and lower joint surfaces was proposed [3]. Results of such algorithm for joint in Fig.1a are presented in Fig.1b and Fig.1c for upper and lower surface, respectively.

In this paper, a shape description methodology, introduced by Jakubowski [6] is applied to hand radiographs interpretation, in order to recognize bones borders shapes in the fingers regions.

2 Analysis of the Characterological Description

Let us recall a formalism presented in [6, 7], where basic unit of the analyzed pattern is one of the sixteen primitives, being line segments or quarters of a circle. Since circle quarters are insufficient to approximate human bones borders, ellipse quarters have been introduced instead of circle ones. Let us also recall definition of a contour $k = p_1 \odot p_2 \odot ... \odot p_m$, where $p_1, p_2, ..., p_m$ are successive primitives of the contour k. Symbols $p_i \odot p_{i+1}$ denotes that p_i is connected to p_{i+1}, such that $hd(p_i) = tl(p_{i+1})$, where $hd(p_k)$ and $tl(p_k)$ corresponds to head and tail of the primitive p_k. Characterological description of contour k contains chain of primitives types and is defined as $char(k) = s_1 s_2 ... s_m$.

To apply recalled formalism, which was originally introduced for the borders of regular objects (e.g. machine parts), let us transform border in Fig.1b into contour in Fig.1d, composed of primitives divided by dots. Then, characterological description of the contour equals:

$$char(k_U) = s_{23}s_{23}s_{43}s_{33}s_{44}s_{24}s_{31}s_{21}s_{41}s_{24}s_{34}s_{31}s_{31}s_{22}s_{22}. \quad (1)$$

On the other hand, characterological description of the contour in Fig.1e, created on the basis of the border in Fig.1c equals:

$$char(k_D) = s_{22}s_{22}s_{41}s_{31}s_{32}s_{41}s_{21}s_{41}s_{44}s_{24}s_{24}s_{24}s_{24}s_{23}s_{33}. \tag{2}$$

Due to the anatomy of the metacarpophalangeal joint, borders presented in Fig.1b and Fig.1c are representative for upper and lower surfaces, respectively. Therefore, their characterological description can be used as patterns in the analysis of other characterological descriptions. This analysis is divided into two steps. In the first step, characterological description is transformed into a chain of biquads (see [6]) using appropriate transducer. In the second step, a finite state machine is used to parse this chain.

2.1 Transducer

According to definition 10, paper [6], transducer is a 5-tuple:

$$T = (G, \Sigma, \Delta, \delta, G_0), \tag{3}$$

where G is a finite nonempty set of states, Σ is a finite nonempty input alphabet, Δ is a finite nonempty output alphabet, G_0 is a finite nonempty set of start states, $G_0 \subset G$ and δ is a finite subset of $G \times \Sigma^* \times \Delta^* \times G$.

Intuitively, if $(q, u, v, q') \in \delta$, it means that the machine is in the state q and the string $u \in \Sigma^*$ is given as an input. Then, the state of the machine is changed into the state q' and $v \in \Delta^*$ becomes the machine output.

Consider the transducer $T_m = (\{q_1, q_2, q_3, q_4\}, S, \{1, 2, 3, 4\}, \delta, \{q_1, q_2, q_3, q_4\})$, where δ is given by the graph depicted in paper [6], Fig.14b. If u causes the transition from the state q_i to q_j, $i \neq j$, then u designates the switch of an (i, j)-biquad (see [6, 7]), what simply means, that there is a switch between i^{th} and j^{th} quadrant.

2.2 Regular Grammars for Chains of Biquads

Let us consider upper surface contour of the healthy joint, presented in Fig.1d. It can be easily noticed, that after proceeding its characterological description with transducer T_m, the following chain is obtained:

$$c_U = 34.41.14.41.12. \tag{4}$$

Moreover, generally chains of the healthy upper surfaces can be described by similar regular expression. As there is regular expression, it is possible to create regular grammar and a corresponding finite state machine to parse chains of other upper surfaces.

In the case of the lower surface, regular expression is more complex and has the following form:

$$(21 + (21.12.21))14(43 + (43.34.43)), \tag{5}$$

whereas, chain obtained for its characterological description equals:

$$c_D = 21.14.43.34.43. \tag{6}$$

In the case of surfaces affected by disease, output chain produced by transducer, usually does not belong to the language corresponding to presented regular expressions. Due to this fact, those chains are not accepted by the finite state machine and recognized as unhealthy. The examples of such kind of borders are presented in Fig.2a and Fig.2b, for upper and lower surfaces, respectively. In such cases, chain produced for upper surface equals 34.41.14.41.14.41.12, whereas chain produced by lower surface equals 21.14.41.14.43.34.43. Obviously, both chains are not accepted by corresponding finite state machines.

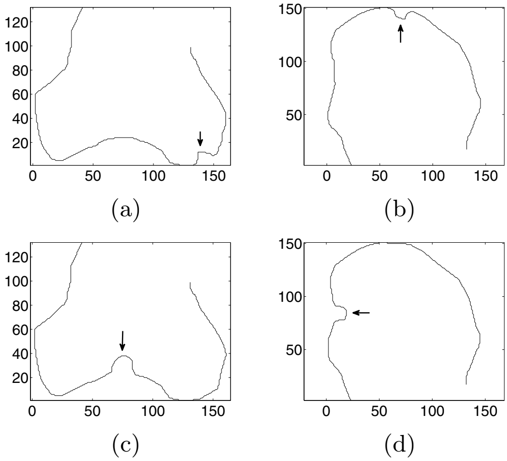


Fig. 2. Borders of upper and lower surfaces affected by disease: lateral erosion of the upper (a) and lower (d) surfaces and central erosion of the upper (c) and lower (b) surfaces

2.3 Lack of the Description Based on Regular Grammars

It is usually possible to separate healthy surfaces from surfaces affected by diseases using regular grammars. However, in some cases this is impossible, due to the general resemblance between healthy and affected bones.

For upper surfaces, this problem is reflected in case of central erosion. As can be observed in Fig.2c, erosion was created around the centre of the bone and thus was merged with the normal shape. As the result, chain produced by transducer will be equal to 34.41.14.41.12 and thus will be accepted by finite state machine of healthy surfaces, although it should be rejected.

Similar problem appears in case of lower surface, presented in Fig.2d, due to the lateral erosion. Obtained chain equals 21.14.43.34.43, just like in case of normal description. Thence detection of the disease is impossible.

To recapitulate, if the investigated characterological descriptions of the contour was not accepted by the transducer, it means that border presents pathological changes. Otherwise, if the characterological descriptions of the contour has been accepted, this means that either the border represents a healthy joint or it represents central erosion in the case of upper surface or lateral erosion in the case of lower surface.

3 Fuzzy Characterological Description of a Contour

Due to the problems related to cases presented in subsection 2.3, the simple analysis, carried out with transducer, is not sufficient in case of some pathological bone changes. Therefore fuzzy characterological description is proposed, to create more detailed description and thus make analysis more accurate.

Algorithm presented in this section, produces fuzzy description of a contour from the list of border points. Thus, it is assumed that there are no redundant points and that the successive points of the border are connected to each other with 8-neighbourhood. That kind of input can be achieved by using algorithms described by authors in [1, 2, 4, 15, 16].

3.1 Fuzzy Primitives

Each of the sixteen defined primitives will be regarded as a fuzzy set with two membership functions. It is due to the fact that bones, as opposed to parts of machinery, slightly vary in shape for different patients. In accordance with this fact and following the terminology proposed by authors in [2], those two membership functions will be considered for successive fragments of border.

The first membership function (μ) depends on the angle between a given primitive and the considered border fragment.

The second membership function (η) is related to the accuracy of the border fragment computed by its comparison to discrete representation of the given primitive (line segment or ellipse quarter) and is also a trapezoid function. Membership functions η corresponds to distance from analyzed border fragment to primitive.

Fuzzy characterological descriptions of a contour is denoted as [[...]...[...]] and consists fuzzy descriptions of the successive border fragments. These, however, are build from primitives descriptors $s_{i,j},\ i, j \in \{1, 2, 3, 4\}$ and values of the membership functions. For example, first fragment in Fig.1b can be described as $[(s_{14}, 0.6, 0.9)(s_{23}, 0.4, 0.9)(s_{33}, 0.4, 0.1)(s_{43}, 0.4, 0.1)]$. Due to the fact that considered border fragment lies in third quarter, only fuzzy primitives from this quarter have membership function μ greater than 0 and therefore the rest of the primitives were omitted.

That kind of description will be investigated to separate healthy joints from joints with central erosion in the case of upper surface or lateral erosion in the case of lower surface. It will be possible due to the additional information - about the angle and the accuracy of the border fragment.

3.2 Border Division

Let *border* $= [P_1 P_2 ... P_n]$ be array with successive points of border going from left to right. The aim of this step is to divide *border* into fragments separated by dots in Fig.1b and 1c. It can be done by direct analysis of the border points coordinates or by using one of the curvature estimators, such as in paper [8].

3.3 Membership Function Depended on Angle

Let $D_1 = [x_{D_1}, y_{D_1}]$ and $D_2 = [x_{D_2}, y_{D_2}]$ be the first and the last points of the considered fragment, respectively and let vector $D_1 D_2$ be defined as follows $D_1 D_2 = [x_{D_1 D_2}, y_{D_1 D_2}] = [x_{D_2} - x_{D_1}, y_{D_2} - y_{D_1}]$. To obtain μ function, an angle between vector and x axis is computed, using the following equation:

$$\alpha = \begin{cases} arccos(\frac{x_{D_1 D_2}}{|D_1 D_2|}), y_{D_1 D_2} \geq 0 \\ 2\pi - arccos(\frac{x_{D_1 D_2}}{|D_1 D_2|}), y_{D_1 D_2} < 0, \end{cases} \tag{7}$$

where $|D_1 D_2|$ is length of vector $D_1 D_2$. Then, value of the membership function μ for each primitive is computed using α angle. If angle is smaller or equal $\frac{\pi}{8}$, membership function μ equals 1. In case of angle from the interval $(\frac{\pi}{8}, \frac{\pi}{4})$, membership function decreases from 1 to 0. In other cases, it equals 0.

3.4 Accuracy Membership Function

The second membership function, η, describes the distance between analyzed fragment, which starts in point D_1 and ends in point D_2 and ideal discrete primitive, which starts and ends in the same points. However, to eliminate primitives from quarter different than this in which the analyzed fragment lies, only those with membership function $\mu > 0$ are computed. If, for some primitive $\mu = 0$, then its second membership function $\eta = 0$ as well. Moreover, discrete representation have to be as close to constant representation as possible and has to be deterministic.

All sixteen primitives have been divided into two groups. The first one contains axial and sloped primitives, which are line segments ($G_1 = \{s_{ij} : i = \{1, 2\}, j = \{1, 2, 3, 4\}\}$). Second group contains concave and convex primitives, which are quarters of an ellipse - not only circles like in case of shape languages ($G_2 = \{s_{ij} : i = \{3, 4\}, j = \{1, 2, 3, 4\}\}$).

Discrete representation of the primitives from first group is computed using Bresenham line algorithm described in [5]. The Bresenham line algorithm determines which points in an image should be plotted in order to form a close approximation to a straight line between two given points. Exampled discrete line segment created using this algorithm for line segment $D_1 D_2$, where $D_1 = [0, 0]$ and $D_2 = [9, 4]$, can be observed in Fig.3 - green line presents constant representation of the line segment; black dots presents its discrete representation.

Another group of primitives requires more complicated algorithm to calculate its discrete representations. In this case, an efficient ellipse-drawing algorithm

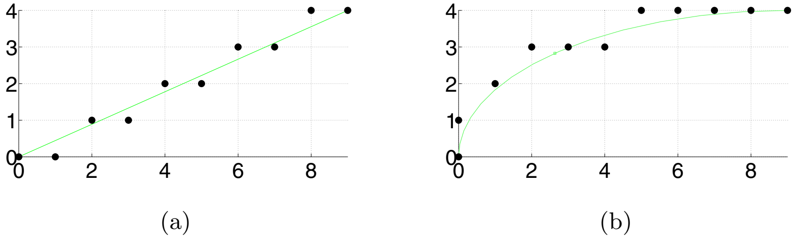


Fig. 3. Constant (line) and discrete (dots) representation of line segment (a) and ellipse (b)

described in paper [14] was used. This algorithm, called also midpoint algorithm, draws ellipses on raster graphics displays. The most important advantage of this algorithm is that it requires only a few integer additions per pixel, so it is very effective. The difference between existing algorithm and algorithm used in this paper is that authors used only quarters of ellipse instead of the whole one. Quarter of the constant and discrete representation of ellipse, corresponding to primitive of type s_{41}, described by the equation $\frac{(x-9)^2}{9^2} + \frac{y^2}{4^2} = 1$ is shown in Fig.3b.

Let $I = [I_1...I_m]$ and $A = [A_1...A_n]$ be successive points of discrete representation of primitive and analyzed fragment, respectively. Let $d^I_{A_i}$, for $i = \{1, ..., m\}$ be minimal distance between point A_i and curve I. Distance between cure I and A is then defined as $D^I_A = \frac{\sum_{i=1}^{m} d^I_{A_i}}{m}$.

As distance correspond to membership function η, it is then defined as value of a trapezoid function with D^I_A as its argument.

At the end of this stage each analysed border fragment is described by chain of triples containing primitive type and two membership functions values $u_i = [(s^i_{i_1 j_1}, \mu^i_1, \eta^i_1)...(s^i_{i_k j_k}, \mu^i_k, \eta^i_k)]$. The characterological descriptions of a contour is then defined as $char(k) = u_1...u_n$.

4 Concluding Remarks

As it was already mentioned, this paper is a continuation of studies described in the series of articles [1, 2, 4, 15, 16]. Summing up, the following has been already done: preprocessing of the whole image, the image segmentation, skeletization, joint localization and surface border extraction. In this paper, a bones characterological descriptions of a contour is obtained and implementation of transducer for border shape analysis is applied. Furthermore, it has been shown that such approach, based on shape description methodology introduced by Jakubowski [6, 7] is sufficient only for preliminary bone border analysis. Therefore, the method improvement, consists in fuzzy sets application, is considered by introducing fuzzy characterological descriptions of a contour with more detailed description than regular characterological descriptions. Fuzzy characterological descriptions parsing is currently a topic of authors investigations and has been generally described in [2]. Our goal is not only to recognize the image, but also to understad it [12] through the use of as much features as possible - including those in papers [1–4, 9, 13, 15, 16].

References

1. Bielecka, M., Skomorowski, M., Bielecki, A.: Fuzzy syntactic approach to pattern recognition and scene analysis. In: Intelligent Control Systems and Optimization, Robotics and Automation, vol. 1, pp. 29–35 (2007)
2. Bielecka, M., Skomorowski, M., Zieliński, B.: A fuzzy shape descriptor and inference by fuzzy relaxation with application to description of bones contours at hand radiographs. LNCS, vol. 5495, pp. 469–478. Springer, Heidelberg (2009)
3. Bielecki, A., Korkosz, M., Wojciechowski, W., Zieliński, B.: Identifying the Borders of the Upper and Lower Metacarpophalangeal Joints Surfaces on Hand Radiographs. In: ICAISC 2010. LNCS (LNAI). Springer, Heidelberg (2010) (accepted)
4. Bielecki, A., Korkosz, M., Zieliński, B.: Hand radiographs preprocessing, image representation in the finger regions and joint space width measurements for image interpretation. Pattern Recognition 41(12), 3786–3798 (2008)
5. Bresenham, J.E.: Algorithm for computer control of a digital plotter. IBM Systems Journal 4(1), 25–30 (1965)
6. Jakubowski, R.: Extraction of Shape Features for Syntactic Recognition of Mechanical Parts. IEEE Transactions on Systems, Man and Cybernetics 15(5), 642–651 (1985)
7. Jakubowski, R., Bielecki, A., Chmielnicki, W.: Data structure for storing drawing being then analysed for purposes of CAD. Archiwa Informatyki Teoretycznej i Stosowanej 1, 51–70 (1993)
8. Kerautret, B., Lachaud, J.O.: Curvature estimation along noisy digital contours by approximate global optimization. Pattern Recognition 42, 2265–2278 (2009)
9. Ogiela, M.R., Tadeusiewicz, R., Ogiela, L.: Image languages in intelligent radiological palm diagnostics. Pattern Recognition 39, 2157–2165 (2006)
10. Sharp, J., Gardner, J., Bennett, E.: Computer-based methods for measuring joint space and estimating erosion volume in the finger and wrist joints of patients with rheumatoid arthritis. Arthritis & Rheumatism 43(6), 1378–1386 (2000)
11. Staniszewska-Varga, J., Szymańska-Jagiełło, W., Luft, S., Korkosz, M.: Rheumatic diseases atlas. Medycyna Praktyczna, Krak"ow (2003) (in Polish)
12. Tadeusiewicz, R., Ogiela, M.R.: Medical Image Understanding Technology. Studies in Fuzziness and Soft Computing. Springer, Heidelberg (2004)
13. Tadeusiewicz, R., Ogiela, M.R.: Picture languages in automatic radiological palm interpretation. International Journal of Applied Mathematics and Computer Science 15(2), 305–312 (2005)
14. van Aken, J.R.: An Efficient Ellipse-Drawing Algorithm. I.E.E.E. Computer Graphics & Applications 4(9), 24–35 (1984)
15. Zieliński, B.: A Fully-Automated Algorithm Dedicated to Computing Metacarpophalangeal and Interphalangeal Joint Cavity Widths. Schedae Informaticae 16, 47–67 (2007)
16. Zieliński, B.: Hand radiograph analysis and joint space location improvement for image interpretation. Schedae Informaticae 17/18, 45–61 (2009)

Localisation and Tracking of an Airport's Approach Lighting System

Shyama Prosad Chowdhury[1], Karen Rafferty[1], and Amit Kumar Das[2]

[1] School of EEECS, Queen's University Belfast, UK
[2] CST Dept., Bengal Engineering and Science University, Shibpur, India
schowdhury01@qub.ac.uk, k.rafferty@ee.qub.ac.uk, amit@cs.becs.ac.in

Abstract. In this paper, we develop novel methods for extracting and tracking regions of interest from a given set of images. In particular, it is our aim to extract information about luminaires making up an airport landing lighting pattern in order to assess their performance. Initially to localise the luminaires we utilise sub pixel information to accurately locate the luminaire edges. Once the luminaires are located within the data, they are then tracked. We propose a new tracking algorithm based on control points and building blocks. Tests performed on a set of 422 images taken during an approach to an airport in Northern Ireland have shown that when combined the localisation and tracking techniques are very effective when compared to standard techniques (KLT and SIFT) as well as to model based matching technique for this application.

Keywords: Photometrics, vibration, luminaire localisation, tracking.

1 Introduction

The airport landing lighting at an airport is used to provide visual cues to a pilot regarding the position and direction of the runway when approaching a given airport. It is important that all the luminaires in the pattern perform according to the standards set by the aviation governing bodies as described in [8]. To date, no physical system exists which can assess the performance of the complete lighting pattern. We propose using a remote camera based assessment technique to solve this problem. A camera is placed inside an aircraft and used to record images of the luminaires during an approach. These images can then be post processed to determine a performance metric for the lighting pattern.

In this paper, we aim to provide a solution to the problem of localising and tracking of luminaires within the image data. Section 3 describes the localisation method of the luminaires using sub pixel analysis. Performance assessment of the luminaires is highly dependent on the accurate tracking of each luminaire. Due to the vibration in aircraft images may become blurred which makes tracking difficult. A novel vibration correction method is described in section 4. Finally the method of tracking each luminaire in the image sequence is outlined in section 5. Note, this paper does not aim to document the techniques for performance assessment using the extracted information.

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 19–26, 2010.

2 Related Research

Segmentation is mainly conducted using the shape, texture or colour information for an object. Saha *et al.* [9] proposes a convex hull based technique to segment out prominent objects from a scene. They have considered that there should be one prominent object in the scene. However, in this research there are multiple objects of interest per image. Nhat *et al.* proposed a different segmentation technique in [3] using a combinatorial graph. This works well for segmenting overlapped objects where the prominent edge information of the objects is present except in the overlapped region. In this research there is no such prominent edge information on overlapping objects.

Because of the large number of luminaires in the lighting pattern and the distance between this and the camera, all the luminaires may appear in a small region of the image and it is very difficult to uniquely identify each luminaire. Any incidence of misinterpretation yields a high negative impact on the accuracy of the performance assessment [1]. Thus prediction of the position of luminaires in the image based on computer vision [5] or random sample consensus (RANSAC) based tracking does not suit properly. Niblock *et al.* used model based tracking [6] to identify luminaires in the collected images. One of the limitations of this work is that accuracy of the tracking is dependant on image quality and it is very difficult to isolate luminaires which are very close together in the image data. Existing known tracking techniques within the community, the Kanade−−Lucus−−Tomasi (KLT) and scale invariant feature transform (SIFT) technique also show very poor success rate [4]. KLT performs at $\sim 60\%$ accuracy whereas SIFT shows only $\sim 20\%$ success rate where,

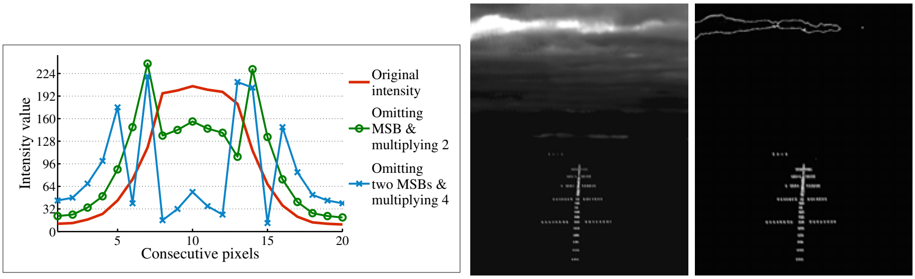
$$\text{Success rate}(\%) = \frac{\text{number of correctly identified luminaires}}{\text{number of luminaires in the image}} 100 \; . \qquad (1)$$

This illustrates the need to develop other tracking techniques for this application.

3 Sub Pixel Luminaire Analysis

Widely used and primitive techniques for localising bright objects from a monochrome image normally utilises one or more threshold values [2], [7], [10]. When such techniques were applied to the collected images runway markings, the sky and luminaires were all identified. Therefore we developed a localisation method based on the sub-pixel data. This is now discussed.

i) Iso-illuminated Ridge (IIR) Image Construction: Each pixel in an image stores a number which relates to the colour of that pixel. For a single byte camera, each pixel can take a value ranging from 0 (black) to 255 (pure white). To date, segmentation techniques have utilised this single pixel value. However, each byte of information is made up from 8 single bits, each of which hold information regarding that pixel. The luminaire is circular in nature and in the centre it normally has a high intensity which decreases towards the perimeter.



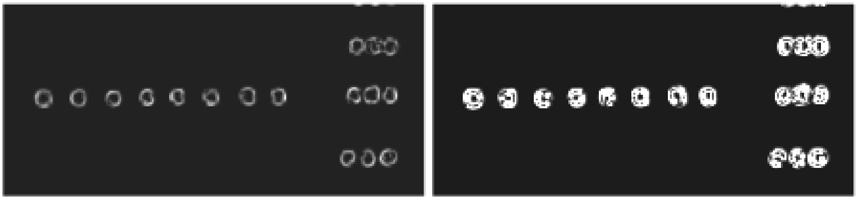
(a) (b) (c)

Fig. 1. (a) Changes in the intensity for different bit movements (b) Original image of the lighting pattern (c) 7-IIRL image

A cross section of the intensity profile along the diameter of a luminaire results in a bell shaped curve (Fig. 1(a)). In a byte, 8 bits (0 to 7) have their different positional weightage factor. Typically the bit at location 7, is termed the most significant bit (MSB). And the bit at location 0, is termed the least significant bit (LSB).By omitting the MSB the highest value of that byte will decreased to 127. To keep the highest value in the same range, the value is multiplied by a factor 2. Similarly after omitting the two MSBs, the value is multiplied by a factor 2^2. After the omission of the MSB and the two MSBs changes in the bell shaped curve is demonstrated in Fig. 1(a) which show a significant edge profile for the luminaire, making it easier to detect. In general α-IIR signifies the omission of all the bits in position same or higher than α followed by a multiplication with $2^{8-\alpha}$. Let the function $b^I(x)$ give the value of the x^{th} bit of the intensity value I then in α-IIR image, the modified value of I will be

$$I_\alpha = \sum_{x=0}^{\alpha-1} b^I(x) 2^{x+1} \ . \tag{2}$$

ii) Iso-illuminated Ridge Line (IIRL) Marking: When the ridges are constructed, a discontinuity in the ridge can be associated with the edge of a luminaire. Fig. 1(b) shows a image taken from the aircraft during landing and



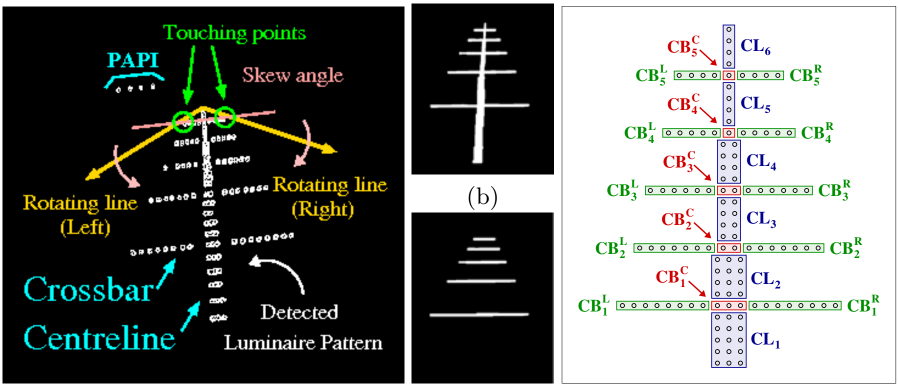
(a) (b)

Fig. 2. (a) Part of the pattern in 7-IIRL image (b) Part of the pattern in 6-IIRL image

Fig. 1(c) illustrates the results when 7-IIRL is applied to that image, with a "close up" given in Fig. 2(a). For comparison the results of 7-IIRL for the same section is shown in Fig. 2(b). By using either 7-IIRL or 6-IIRL only the luminaires within the image are localised. This is a very positive result. Indeed this technique would be applicable in the other research domain as a reliable means of detecting light sources in image data.

4 Skew Correction and Region of Interest

Having identified each luminaire with an IIRL, it is now necessary to correlate each luminaire with a physical description of the lighting pattern in order to uniquely identify each luminaire. However because of aircraft movement, images of the lighting are frequently rotated in the 2D image plane of the camera. In an ideal situation, the centreline and crossbars (Fig. 3(a)) should appear in the vertical and horizontal directions of the image plane. Typically however they appear skewed within the image. It is necessary to remove the skew to simplify luminaire tracking. To do this, we propose a two step technique.



(a) (c) (d)

Fig. 3. (a) Major skew correction (b) Composite structure after closing (c) Detected crossbars after opening (d) Control blocks formation using the lighting pattern

4.1 Major Skew Correction (MSC)

In order to quickly identify any major skew within the image of the lighting it is necessary to define control points (CP) in the pattern. The top luminaire in the centreline is automatically chosen as CP-1. From CP-1 two straight lines are formed on both the left and right side of the pattern. Initially the two lines are projected horizontally in the image. The left line is rotated anticlockwise until an intersection occurs between it and the IIRL. A similar process is carried out for the right line except it is rotated in the clockwise direction. Most likely this intersection will occur at the top crossbar (see Fig. 3(a)), however this can not be guaranteed in the case of missing luminaires. In that situation using the nearest intersected point another end is obtained. Let the final touch points on the same crossbar be the P^l and P^r. For any point x, the function $MVP(x)$ finds the middle point of the vertical cross section of the component. The angle function $AN(x, y)$ finds the angle in between two points x and y. Thus the angle between the points $MVP(P^l)$ and $MVP(P^r)$ are measured in β^M,

$$\beta^M = AN(MVP(P^l), MVP(P^r)) \; . \tag{3}$$

The image is then rotated in $-\beta^M$ direction to correct the major skew. Using this technique most of the skew can be removed from the image.

4.2 Finer Skew Correction (FSC)

FSC is achieved by measuring the deflection of the crossbars from the horizontal direction. In order to determine the crossbars, morphological filters are utilised. Morphological operations are undertaken using a variable length structuring element (VLSE).

i) Composite Structure Construction: At any row position i, the function $ASC(i)$ gives the average horizontal spread of the centreline. The approximated row position of the first and second crossbars are labelled as R_1 and R_2 if they are present in the image. Let $VLSE_C$ be used for morphological closing where, the height and width of this VLSE are denoted by $VLSE_C^H$ and $VLSE_C^W$. Both $VLSE_C^H$ and $VLSE_C^W$ are two dimensional image functions where, $VLSE_C^H(i, j)$ and $VLSE_C^W(i, j)$ represents the height and width of the VLSE on a pixel in i^{th} row and j^{th} column. Here $VLSE_C^H(i, j) = \lambda_1 ASC(i)$ and $VLSE_C^W(i, j) = \lambda_2 ASC(i)$ where, the value of λ_1 is kept 2 and λ_2 is 3. The reason for choosing these particular λ values is to ensure that the morphological closing operation should not miss any present luminaires. Morphological closing on an IIRL image with $VLSE_C$ results in a single structure with all the luminaires. This image will be known as $IIRL^C$ for future use. Fig. 3(b) shows the composite structure after the morphological closing operation.

ii) Crossbars Extraction: Let $VLSE_O$ be used for morphological opening where, height and width of this VLSE are denoted by $VLSE_O^H$ and $VLSE_O^W$. Opening is performed with a horizontal line like structuring element (SE). Thus, $VLSE_O^H$ is of a constant value 1 and $VLSE_O^W(i, j) = \lambda ASC(i)$. The value of λ is kept as 2 to ensure that the opening element will definitely remove the centreline links in between two crossbars to form them like horizontal lines. Fig. 3(c) illustrates the results after VLSE.

iii) Deflection Angle Calculation: Connected component analysis (CCA) on the residual after binary morphology opening will label all the crossbars. Let us assume a total T $(1 \le T \le 5)$ number of components are present where the spread of the k^{th} component is from the point P_k^l to P_k^r. The distance function $DS(x, y)$ determines the distance in between two points x and y. As defined earlier, the angle function $AN(x, y)$ measures the angle. Average deflection of all the crossbars from horizontal line is calculated as β^F where,

$$\beta^F = \sum_{k=1}^{T} DS(P_k^l, P_k^r) AN(P_k^l, P_k^r) \ / \ \sum_{k=1}^{T} DS(P_k^l, P_k^r) \ . \tag{4}$$

The $IIRL^C$ image is rotated with the $-\beta^F$ angle to get the $IIRL^{CR}$ which is fully skew corrected ($< 0.5°$) and all the crossbars become horizontally aligned.

Finally having identified each crossbar and the centreline a number of control blocks can be automatically defined using segmentation techniques. The final defined 21 control blocks are shown in Fig. 3(d).

5 Pattern Based Tracking of ALS Luminaires

At this stage a different tracking scheme is applied to the centreline (CL_i) , wing $((CB_i^L)$, $(CB_i^R))$ and centre body blocks $((CB_i^C))$.

i) Tracking the Centreline Blocks: In each of the centreline blocks, binary morphological closing is done with a VLSE $(VLSE_O)$. For a given block if the expected number of luminaires are not found, then the centreline block which is in between two crossbars CB_k and CB_{k+1} are vertically divided into four regions. However, the division of the regions can not be linear because of the perspective error. Using all the vertical distances in between crossbars, amount of perspective error is calculated and then corrected. Now the sets in a block (k) is horizontally analysed in the same method as described for the centre body blocks tracking. This uses the virtual central line (VCL_k) which is defined as the the linear regression calculated on all the luminaires present in the block CB_k.

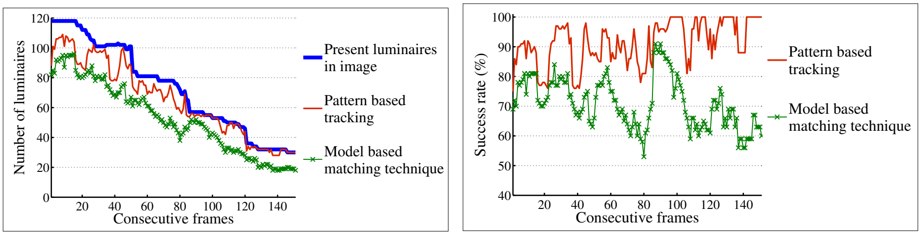
ii) Tracking for Centre Body Blocks: Use of the VCL_k is used to analyse the $(k-1)^{th}$ centre body (CB_{k-1}^C). If there are three luminaires in the centre body block's pattern then VCL_k is used to track the position of the middle one and the position of the others two can be determined from that. Where two luminaires are in the centre body block, VCL_k will act as a clear separator among them.

iii) Tracking for Wing Blocks: It is observed from the ALS pattern that the intermediate distance among the luminaires in the wing blocks are highest compare to any other blocks. Both of the left and right wings are analysed together to find the median distance among the horizontally distributed luminaires. Using this distance and the absence history all the luminaires in the wing are tracked.

6 Results

The localisation and tracking algorithms were applied to two sets of videos taken during an approach to an airport in Northern Ireland. Each video had approximately 200 images. Thus over 400 images were tested using the new algorithms.

Initially when the IIR analysis was applied to the images, it was shown that removal of the two MSBs from the image data, could effectively create a large intensity shift at the edges of the luminaire, aiding in its detection. It was found that this technique was also very effective and its accuracy is 100% for localising all the prominent luminaires. In addition, a noticeable benefit to using the IIR analysis, is that it was also very effective for isolating very small luminaires in the image data. For example, a number of luminaires at the top of the pattern only cover 7 pixels at the start of the approach. This outperforms other published techniques for this applications [6].



(a) (b)

Fig. 4. (a) Number of luminaires per image, the number of luminaires correctly tracked by the Pattern based tracking and the Model based matching technique (b) Success rate (%) of the correctly tracked luminaires

Having successfully located the luminaires within the image data, the developed tracking technique was applied to the images. Success rate on the tracking is measured in each frame by the ratio of the correctly tracked luminaires and total number of luminaires visible in the image (Eq. 1). It has already reported in [4] that the existing KLT and SIFT technique gives $< 60\%$ and $< 20\%$ accuracy in tracking of the luminaires and model based matching technique produces much better result compare to the others. So, we only compared our result (pattern based tracking) with the model based matching (Fig. 4(a) and 4(b)). In our method, when a luminaire is clearly distinguishable within the images, the tracking accuracy is 100%. However, this reduced when the luminaires are highly overlapped. Thus in terms of the tracked luminaires in the video sequence an overall success rate of 91% for tracked luminaires was achieved. This again performs well when compared to other published tracking algorithms in the area [4]. It is also interesting to note that the processing time for the localisation and tracking of luminaires per image was found to take 1 second (for Model based matching technique it is 2.3s/image), when running on a standard pc. Whilst not real time, the processing speed is still very acceptable.

7 Conclusion

The authors present new techniques for localising and tracking luminaires within images. In particular, the techniques have been applied to images of an airport lighting pattern that were collected during an approach to an airport in Northern Ireland. The new technique for localising the luminaires utilises sub pixel information in order to determine the edges of the luminaires. This technique would work well for any application that requires the localisation of a light source within an image. Tracking of a known pattern is discussed and performs very well. Again this tracking technique could be applied to any video sequence where there is a known pattern of points of interest. Finally it can be concluded that the complete preprocessing presented here is very useful for the target application. Future work will concentrate on using the information extracted from the images to determine the performance metric of the lighting.

Acknowledgements. The authors would like to thank the EPSRC (Grant: EP/D05902X/1) for financial backing. The contribution of Flight Precision and Belfast International Airport for providing flight time in order to collect airport lighting data is also gratefully acknowledged.

References

1. Chowdhury, S., McMenemy, K., Peng, J.: Performance evaluation of airport lighting using mobile camera techniques. In: Jiang, X., Petkov, N. (eds.) CAIP 2009. LNCS, vol. 5702, pp. 1171–1178. Springer, Heidelberg (2009)
2. Liu, J.: Robust image segmentation using local median. In: Proc. of the 3rd Canadian Conf. on Computer and Robot Vision, CRV 2006. IEEE Computer Society Press, Los Alamitos (2006)
3. Nhat, V., Manjunath, B.: Shape prior segmentation of multiple objects with graph cuts. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, June 23-28, pp. 1–8 (2008)
4. Niblock, J., Peng, J., McMenemy, K., Irwin, G.: Fast model-based feature matching technique applied to airport lighting. IET Science, Measurement and Technology 2(3), 160–176 (2008)
5. Niblock, J., McMenemy, K., Ferguson, S., Peng, J.: Autonomous tracking system for airport lighting quality control. In: 2nd International Conference on Computer Vision Theory and Applications, March 8-11, vol. (2), pp. 317–324 (2007)
6. Niblock, J., Peng, J., McMenemy, K., Irwin, G.: Autonomous model-based object identification and camera position estimation with application to airport lighting quality control. In: Proc. of 3rd Int. Conference on Computer Vision Theory and Applications, VISAPP, Funchal, Portugal, January 12-15, vol. (2) (2008)
7. Okada, K., Akdemir, U.: Blob segmentation using joint space-intensity likelihood ratio test: application to 3d tumor segmentation. In: IEEE Conf. on Comp. Vision and Pattern Recognition. CVPR 2005, June 20-25, vol. 2, pp. 437–444 (2005)
8. Organization, I.C.A.: Aerodrome design and operations, annex 14, 4th edn., vol. 14 (July 2004)
9. Saha, S., Das, A., Chanda, B.: An automatic image segmentation technique based on pseudo-convex hull. In: Proc. of Indian Conference on Computer Vision, Graphics and Image Processing, ICCVGI 2006 (2006)
10. Wong, W., Chung, A.: Bayesian image segmentation using local iso-intensity structural orientation. IEEE Transaction on Image Processing 14(10), 1512–1523 (2005)

Algorithm for Blood-Vessel Segmentation in 3D Images Based on a Right Generalized Cylinder Model: Application to Carotid Arteries

Leonardo Flórez Valencia[1], Jacques Azencot[2], and Maciej Orkisz[2]

[1] Pontificia Universidad Javeriana
Departamento de Ingeniería de Sistemas, Bogotá, Colombia
florez-l@javeriana.edu.co

[2] Université de Lyon; Université Lyon 1; INSA-Lyon
CNRS UMR5220; INSERM U630; CREATIS, F-69621 Villeurbanne, France
{azencot,orkisz}@creatis.insa-lyon.fr

Abstract. The arterial lumen is modeled by a spatially continuous right generalized cylinder with piece-wise constant parameters. The method is the identifies the parameters of each cylinder piece from a series of planar contours extracted along an approximate axis of the artery. This curve is defined by a minimal path between the artery end-points. The contours are extracted by use of a 2D Fast Marching algorithm. The identification of the axial parameters is based on a geometrical analogy with piece-wise helical curves, while the identification of the surface parameters uses the Fourier series decomposition of the contours. Thus identified parameters are used as observations in a Kalman optimal estimation scheme that manages the spatial consistency from each piece to another. The method was evaluated on 15 datasets from the MICCAI 3D Segmentation in the Clinic Grand Challenge: Carotid Bifurcation Lumen Segmentation and Stenosis Grading (http://cls2009.bigr.nl/). The average Dice similarity score was 71.4.

1 Introduction

The goal of this work is the segmentation of carotid arteries.Complex pathological deformations of normally circular vascular cross-sectional shapes can be represented by the Right Generalized Cylinder state model (RGC-sm) proposed in [1]. The model is composed of two parts respectively describing: 1) the axial shape by means of orthonormal bases attached to the axis, and 2) the surface by means of planar contours orthogonal to the axis. According to the system-state formalism, both components (the local base and the cross-sectional contour) corresponding to any arc-length location, can be calculated knowing only one initial base and contour, as well as their dynamics (parameters describing their variation). Our algorithm reconstructs the RGC shape step-by-step, by extracting consecutive cross-sectional contours along an approximate vessel axis. At each step, the RGC parameters are calculated from the current and previous contour. The continuity and smoothness are controlled by a Kalman

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 27–34, 2010.

state estimator (KSE) [2]. A preliminary version of this work is available on-line (`http://hdl.handle.net/10380/3106`).

2 RGC Model Overview

The RGC-sm associates a generating curve $\mathcal{H}$ and the surface $\mathcal{S}$. It assumes that a cylinder can be subdivided into pieces such that the model parameters be piece-wise constant. Each piece $\mathcal{H}_i$ of the generating curve is defined by its length $\varDelta_i$, curvature κ_i, torsion τ_i and by the azimuthal rotation angle ν_i of the local base $\boldsymbol{\varGamma}_i(t)$ attached to $\mathcal{H}_i$, with respect to the corresponding Frenet frame. Each piece $\mathcal{S}_i$ of the surface, *i.e.* a continuous stack of contours $\mathbf{c}_i(t,\omega)$, is defined by the Fourier coefficients $\mathbf{Z}_i = \{z_{i,l} \in \mathbb{C}; -q \leq l \leq +q\}$ describing the first contour in the piece, and by $\boldsymbol{\varLambda}_i = \{\lambda_{i,l} \in \mathbb{C}; -q \leq l \leq +q\}$, an ordered set of $2q+1$ coefficients linearly transforming the contour along $\mathcal{H}_i$:

$$\mathbf{c}_i(t,\omega) = \sum_{l=-q}^{+q} (\lambda_{i,l}(t-t_i) + z_{i,l})e^{jl\omega}, \tag{1}$$

where t and ω respectively are arc-length and azimuthal parameters. The number q of harmonics controls the level of details. Each surface piece $\mathcal{S}_i$ is connected to the corresponding generating curve piece $\mathcal{H}_i$ by the following equation:

$$\mathbf{s}_i(t,\omega) = \boldsymbol{\varGamma}_i(t) \cdot \begin{bmatrix} 0 \\ \mathtt{Re}(\mathbf{c}_i(t,\omega)) \\ \mathtt{Im}(\mathbf{c}_i(t,\omega)) \end{bmatrix} + \mathbf{h}_i(t), \tag{2}$$

where $\mathbf{h}_i(t)$ is the spatial location of the origin of $\boldsymbol{\varGamma}_i(t)$, which belongs to $\mathcal{H}_i$. The entire model is thus $\mathcal{M} \equiv \{\mathbf{h}_0, \boldsymbol{\varGamma}_0, \mathbf{Z}_0, \{\kappa_i, \tau_i, \nu_i, \varDelta_i, \boldsymbol{\varLambda}_i;\ 0 \leq i < n\}\}$,where $\mathbf{h}_0 \equiv \mathbf{h}_0(t=0)$ is the first point of $\mathcal{H}$, $\boldsymbol{\varGamma}_0 \equiv \boldsymbol{\varGamma}_0(t=0)$ is the first base attached to $\mathbf{h}_0$ and $\mathbf{Z}_0 \equiv \mathbf{Z}_0(t=0)$ is the Fourier decomposition of the first contour.

3 Retrieval of Model Parameters from Image Data

The vascular segment of interest is delimited by two points interactively given by the user. The initialization of the RGC-sm reconstruction process requires $\mathbf{h}_0$, $\boldsymbol{\varGamma}_0$ and $\mathbf{Z}_0$. One of the seed-points (typically the distal one) is taken as $\mathbf{h}_0$. The extraction of the first discrete contour $\mathcal{C}_0 = \{\mathbf{c}_{0,k} : 0 \leq k < K\}$ is initialized by $\mathbf{h}_0$ and is performed in the plane orthogonal (in $\mathbf{h}_0$) to the initial approximate axis $\tilde{\mathcal{H}}$. The orthonormal base $\boldsymbol{\varGamma}_0$ is constructed such that its first vector is tangent to $\tilde{\mathcal{H}}$ in $\mathbf{h}_0$, the second vector is oriented along the vector $\mathbf{h}_0 - \mathbf{c}_{0,0}$, and the third one is orthogonal to both. $\mathbf{Z}_0$ is calculated as the Fourier series corresponding to $\mathcal{C}_0$, using the harmonics up to the third order ($q = 3$), which is sufficient to represent reasonably complex cross-sections.

The KSE predicts the 3D locations, orientations and shapes of the contours that delimit the consecutive cylinder pieces [3]. The observation vector, coding the RGC-sm parameters of the i-th piece, is computed from the result of the contour extraction performed in the predicted plane. The observation is then filtered by the KSE to produce a corrected estimate of the parameters, which is expected to smooth out the possible errors of the contours and initial axis extraction. Only the very first contour remains uncorrected. The KSE adapts the tracking speed to the complexity of the local vascular shapes (axial and superficial). In complex shapes (high local changes of curvature, for example) the length of the cylinder piece is automatically reduced. This occurs when the predicted contours are too different from the observations.

Approximate Axial Shape Extraction

The initial line $\tilde{\mathcal{H}}$ coarsely describing the axial shape of the vessel, is constructed as follows:

1. A binary image $\mathcal{B}(\mathbf{p})$ coarsely representing the vascular lumen is computed from the initial image $f(\mathbf{p})$, using a flooding algorithm. This algorithm uses locally adaptive thresholds f_L, f_H depending on the distance of each voxel to the seeds. These local values are computed using the Robust Automatic Threshold Selection (RATS) scheme [4] and the properties of the vascular lumen intensities along the carotids, presented in [5].
2. An Euclidean distance map $\mathcal{E}(\mathbf{p})$ is computed within $\mathcal{B}(\mathbf{p})$ using the algorithm proposed in [6].
3. A cost function $\mathcal{F}(\mathbf{p}) = \frac{1}{1+\mathcal{E}(\mathbf{p})}$ is used for a *minimal path* algorithm (as proposed in [7]), which finds a set $\tilde{\mathcal{H}}$ of points ordered along the shape and connecting the seed-points even in the presence of lumen discontinuities (severe stenoses). These points are expected to be located near the center of the vessel due to the cost function $\mathcal{F}(\mathbf{p})$, related to the distance map $\mathcal{E}(\mathbf{p})$.
4. The curve is smoothed by use of a Laplacian filter.

Contours Extraction

The planar contours are extracted using the *fast marching* (FM) method [8]. FM is a *front propagation* technique that provides a set of (counter-clockwise) *ordered* points $\mathcal{C}_i = \{\mathbf{c}_{i,k} : 0 \leq k < K\}$. This technique needs the definition of a potential field $\mathcal{P}(\mathbf{p})$ expected to be maximum at discontinuities (edges) and minimum within uniform regions. We use the potentials proposed in [9], where the native image intensities $f(\mathbf{p})$ along with image discontinuities represented by the gradient magnitude $|\nabla f(\mathbf{p})|$ are used:

$$\mathcal{P}\left(\mathbf{p}\right) = \left|\nabla f\left(\mathbf{p}\right)\right| \mathcal{K}\left(d\left(\mathbf{p}\right), \alpha_1, \beta_1\right) \mathcal{K}\left(f\left(\mathbf{p}\right), \alpha_2, \beta_2\right) \mathcal{K}\left(f\left(\mathbf{p}\right), \alpha_3, \beta_3\right) \quad . \tag{3}$$

The function $\mathcal{K}(\cdot)$ of a scalar u is defined as follows:

$$\mathcal{K}\left(u, \alpha, \beta\right) = 1 + \exp\left(-\frac{u-\alpha}{\beta}\right) \quad , \tag{4}$$

where α is a shift, β is a signed sharpness parameter and $d(\mathbf{p})$ is a monodimensional derivative of $|\nabla f(\mathbf{p})|$ along the direction normal to the front. With $\alpha_1 = 0$, $\alpha_2 = f_L$, $\alpha_3 = f_H$, $\beta_1 = 100$, $\beta_2 = 1$ and $\beta_3 = -1$, this potential function $\mathcal{P}$ strongly decreases the propagation speed when the front moves beyond local maxima of the gradient norm (i.e. beyond borders of the arteries) or beyond the range $[f_L, f_H]$ of luminal intensities, defined as described in previous subsection. Furthermore, in this work the authors recommend that the FM propagation should be stopped at time value T when the growth of the area A encompassed by the front becomes very slow, which is characterized by a large value of $\varDelta T/\varDelta A$. Actually, at each iteration we compute the mean $m_{T/A}$ and the standard deviation $\sigma_{T/A}$ of $\varDelta T/\varDelta A$ and the propagation stops when $\varDelta T/\varDelta A > m_{T/A} + 10\sigma_{T/A}$. All the coefficient settings correspond to the experimental values from [9].

Additionally, the search space of the FM algorithm is restricted as follows: 1) the distance map values on the initial line $\tilde{\mathcal{H}}$ are low-pass filtered along $\tilde{\mathcal{H}}$, within the binary mask $\mathcal{B}(\mathbf{p})$, 2) in the segments of $\tilde{\mathcal{H}}$ falling beyond $\mathcal{B}(\mathbf{p})$, the distance values are interpolated between the values defined within $\mathcal{B}(\mathbf{p})$, 3) thus calculated distance values are used as an additional stopping criterion.

Front propagation is performed in the plane passing through the predicted point $\hat{\mathbf{h}}_i \in \tilde{\mathcal{H}}$, and oriented according to the predicted vessel orientation expressed by the orthogonal base $\hat{\varGamma}_i$. Summarizing, the point set $\mathcal{C}_i$ is extracted as follows:

1. The potential field $\mathcal{P}(\mathbf{p})$ is sliced by the plane passing by $\hat{\mathbf{h}}_i$ and oriented by the first column vector of $\hat{\varGamma}_i$, to obtain a 2D image $\mathcal{Q}_i(\mathbf{p})$.
2. The FM algorithm is executed on $\mathcal{Q}_i(\mathbf{p})$ with $\hat{\mathbf{h}}_i$ as the first trial point (front initialization), and the propagation stops when $\varDelta T/\varDelta A > m_{T/A} + 10\sigma_{T/A}$ or when the first point of the front reaches the above defined distance from the initialization point.
3. The FM generates the level set $\mathcal{L}_i(\mathbf{p})$ which contains $\mathcal{C}_i$ as its last level.

Calculation of Model Parameters

Without loss of generality, we explain the process for the first cylinder piece. Under the assumption of constant curvature and torsion, each piece of $\mathcal{H}$ is a helix, and a geometrical reasoning demonstrates that its parameters can be recovered if the frames $\boldsymbol{\varGamma}_0$, $\boldsymbol{\varGamma}_1$ at its extremities, as well as their origin locations $\mathbf{h}_0$ and $\mathbf{h}_1$, are available. The remaining parameters are calculated using the Fourier decompositions of the contours $\mathbf{Z}_0$ and $\mathbf{Z}_1$.

We first compute the transition operators, respectively rotation and translation, between the extremities:

$$\boldsymbol{\varPhi}(0, \varDelta_0) = \boldsymbol{\varGamma}_0^\top \cdot \boldsymbol{\varGamma}_1, \tag{5}$$

$$\mathbf{Tr}(0, \varDelta_0) = \boldsymbol{\varGamma}_0^\top \cdot (\mathbf{h}_1 - \mathbf{h}_0). \tag{6}$$

$\boldsymbol{\varPhi}$ is a rotation operator and one of the properties of the rotation matrices is the existence of $\varTheta$ and $\boldsymbol{\varXi}$ such that $\boldsymbol{\varPhi}(0, \varDelta_0) - \boldsymbol{\varPhi}(0, \varDelta_0)^\top = 2 \sin\varTheta \cdot \boldsymbol{\varXi}$, where $\varTheta$

is the rotation angle and the non-zero elements of the antisymmetric matrix $\boldsymbol{\varXi}$ compose the vector defining the rotation axis. Moreover, it can be demonstrated that $\boldsymbol{\varPhi}(t_1, t_2) = \exp((t_2 - t_1) \cdot \boldsymbol{\varPsi})$, where $\boldsymbol{\varPsi}$ is proportional to $\boldsymbol{\varXi}$: $\boldsymbol{\varPsi} = \mu\boldsymbol{\varXi}$. In the case where $\boldsymbol{\varPhi}$ describes the rotation between two frames attached to a helix, the coefficient μ is also involved in the calculation of the curvature and torsion of the helix. Indeed, these are proportional (via μ) to the cosine and sine of the angle φ that represents the "slope" of the helix. Hence, the second step is the subtraction:

$$\boldsymbol{\varPhi}(0, \varDelta_0) - \boldsymbol{\varPhi}(0, \varDelta_0)^\top = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix} \Rightarrow \varTheta = \arcsin\left(\sqrt{a^2 + b^2 + c^2}/2\right).$$

It can be demonstrated that a, b, c can be used to calculate φ, then μ:

$$\varphi = \arcsin\left(a/\sqrt{a^2 + b^2 + c^2}\right), \qquad \mu = \frac{\sin\varTheta\cos^2\varphi + \varTheta\sin^2\varphi}{\left[\boldsymbol{\varPhi}(0, \varDelta_0)\,[1\;0\;0]^\top\right]^\top \mathbf{Tr}(0, \varDelta_0)},$$

which in turn permits the computation of the axial parameters:

$$\begin{cases} \nu_0 &= \arctan(b/c), \\ \kappa_0 &= \mu\cos\varphi, \\ \tau_0 &= \mu\sin\varphi, \\ \varDelta_0 &= \varTheta/\mu. \end{cases} \tag{7}$$

Numerical stability problems might arise when $\boldsymbol{\varPhi}(0, \varDelta_0) = \mathbf{I}$, which occurs when $\mathcal{H}_0$ is a straight line segment. This is checked after the computation of $\boldsymbol{\varGamma}_0$ and $\boldsymbol{\varGamma}_1$, and the parameters, in this case, are set as follows: $\left[\,\kappa_0\;\tau_0\;\nu_0\;\varDelta_0\,\right]^\top = \left[\,0\;0\;0\;|\mathbf{h}_1 - \mathbf{h}_0|\,\right]^\top$. Another important point of the numerical implementation is the appropriate choice between two solutions, theorically equally valid, respectively associated with the rotation $\varTheta$ and the counter-rotation $2\pi - \varTheta$. The algorithm first checks that the resulting $\varDelta$ is positive. Then, it checks that thus estimated end-point $\tilde{\mathbf{h}}_1$ of the resulting helix falls within the corresponding contour, more precisely within its fundamental ellipse. If both solutions meet these criteria, the one that corresponds to the shortest curve is kept, *i.e.* $\min(\varTheta, 2\pi - \varTheta)$ is chosen. Otherwise, the solution with smallest distance between $\tilde{\mathbf{h}}_1$ and the contour center is kept.

The last step is the computation of the parameters describing the linear evolution of the Fourier decomposition of the contours:

$$\boldsymbol{\varLambda}_\mathbf{0} = \left\{\frac{z_{1,l} - z_{0,l}}{\varDelta} \in \mathbb{C} : -q \leq l \leq +q\right\}. \tag{8}$$

4 Results

The method was evaluated on 15 Computed Tomography Angiography (CTA) datasets available within the Carotid Bifurcation Algorithm Evaluation Framework (`http://cls2009.bigr.nl`). These datasets are provided together with reference standard constructed from manual annotations performed by 3 experts per dataset, as well as with three initialization points per artery, located in the common, external and internal carotid arterial segments, *i.e.* CCA, ECA and ICA respectively. Therefore, for each carotid artery we extracted two cylinders, respectively corresponding to CCA+ICA and CCA+ECA. The final result of our segmentation was the union of the two cylinders (overlapping in the CCA part). According to the framework, the evaluation in each dataset was performed on a volume of clinical interest defined as follows: 20mm below from the bifurcation plane in the CCA, 40mm above the bifurcation plane in the ICA and 10mm above the bifurcation plane in the ECA. Four criteria were evaluated to assess the lumen segmentation: Dice similarity index, mean surface distance, RMS surface distance and maximal surface distance. Details of the evaluation methodology can be found in [10].

Figure 1 displays examples of qualitative segmentation results. The quantitative results are provided in Table 1. The average running time of our algorithm, including lumen segmentation and display of intermediate and final results, was less than 2 minutes per artery (ICA+ECA+CCA), which actually means 57 sec. per cylinder (ICA+CCA or ECA+CCA).

5 Discussion and Conclusion

RGC-sm is a powerful tool that permits a concise description of complex generalized cylindrical shapes. The theoretical framework permits the reconstruction of a continuous surface corresponding to the lumen, based on a limited number of discrete contours. Additionally, the Kalman estimator permits a correction of the observation errors when these remain within a reasonable range. However, the evaluation results are not completely satisfactory. One can argue that the overlap between the segmentation result and the reference would be better if the whole artery was considered. Indeed, owing to pathological deformations and image intensity alterations (calcifications, hypodense plaque), the volume of interest used for the evaluation is clearly the most difficult part of the artery. Nevertheless, our current implementation of the image processing steps devised to provide the observations is not optimal. The initial rough extraction of the axial shape begins by a thresholding step, which is prone to errors when neighboring structures have a similar intensity range. Actually, this initial curve needs to be quite well-centered within the lumen. Indeed, as the Fast Marching algorithm in planes orthogonal to this initial curve starts from the intersection between the plane and the curve, this intersection has to fall within the lumen, otherwise the contour extraction fails. Failures are also observed when the curve is located within the lumen, close to a poorly contrasted boundary. In fact, the use of an always inflating deformable contour, such as the Fast Marching front, is uneasy, since the speed function and stopping criteria hardly can cope with all possible configurations (nearby veins, calcifications, etc.). Furthermore, 2D Fast Marching does not exploit the 3D continuity of the vascular lumen, which might be helpful in some complicated situations. Future work will be oriented towards

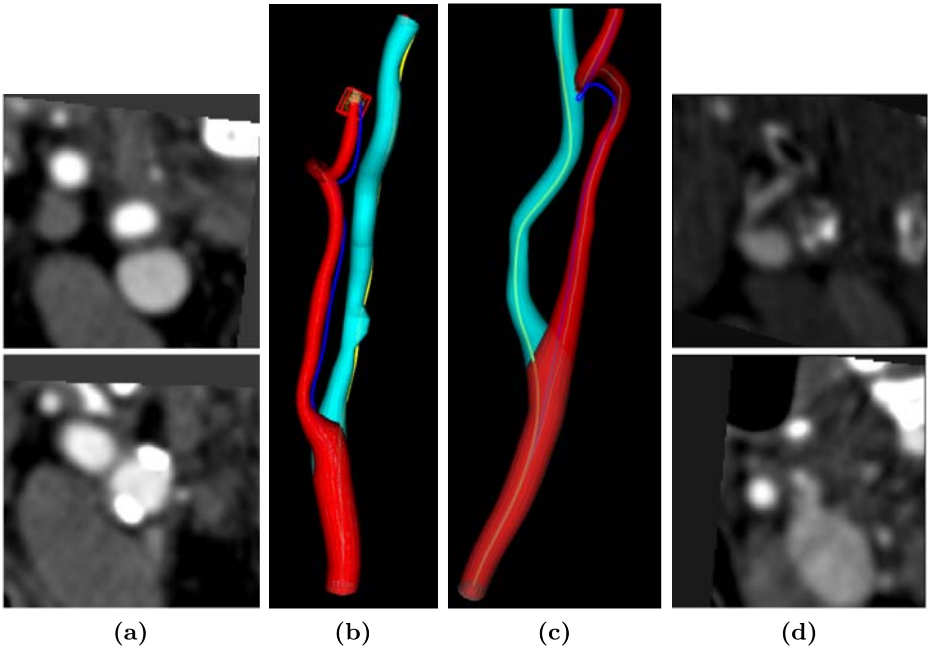


Fig. 1. Two examples of the segmentation of a CTA image of a carotid using our algorithm. The internal carotid is presented in cyan and the external carotid is presented in red. The yellow and blue lines represent the generating curves calculated. Examples of cross-section show various levels of difficulties in boundary detection: (a top) well defined lumen, (a bottom) lumen with 2 bright calcifications, (d top) hardly identifiable lumen surrounded by a heterogeneous plaque, and (d bottom) lumen stuck to a nearby structure of same density. Please note that the artery center is always at the center of the image.

Table 1. Summary of the quantitative evaluation

Measure	**min.**	**max.**	**avg.**
Dice similarity index	32.89%	98.21%	73.16%
mean surface distance	0.55mm	3.52mm	1.58mm
RMS surface distance	0.74mm	5.12mm	2.55mm
maximal surface distance	2.32mm	16.22mm	8.50mm

an implementation that do not require the initial extraction of an approximate axial shape and that perform a piece-wise local 3D boundary extraction. Let us also note that the current implementation was designed with an implicit assumption (constraint) that the seed-points are also end-points, and are given at "easy" locations, *i.e.* healthy circular cross-sections without neighboring structures "stuck" to the artery of interest. Hence, the method had difficulties in datasets where the seed-point in the internal carotid artery was located either close to the bifurcation, so that our method segmented a too short part of the artery, or near the skull where the assumption of the absence of neighboring structures did not hold. Further work is necessary to cope with seed-points located anywhere within the vessel of interest.

Acknowledgements

This work has been supported by the Région Rhône-Alpes (France) via the Simed project of the ISLE research cluster and by the internal project 2077 ”Vitral 3” at the Pontificia Universidad Javeriana.

References

1. Azencot, J., Orkisz, M.: Deterministic and stochastic state model of right generalized cylinder (RGC-sm): application in computer phantoms synthesis. Graph. Models 65, 323–350 (2003)
2. Kalman, R.: A New Approach to Linear Filtering and Prediction Problems. Trans ASME–J. Basic Engineering 82, 35–45 (1960)
3. Flórez Valencia, L., Azencot, J., Vincent, F., Orkisz, M., Magnin, I.: Segmentation and Quantification of Blood Vessels in 3D Images using a Right Generalized Cylinder State Model. In: Proc. IEEE Int. Conf. Image Process., pp. 2441–2444 (2006)
4. Kittler, J., Illingworth, J., Fglein, J.: Threshold selection based on a simple image statistic. Comput. Vision Graphics Image Process. 30, 125–147 (1985)
5. Flórez Valencia, L., Vincent, F., Orkisz, M.: Fast 3D pre-segmentation of arteries in computed tomography angiograms. In: Int. Conf. Comput. Vision & Graphics, Warsaw, Poland, pp. 87–88 (2004)
6. Maurer Jr., C.R., Qi, R., Raghavan, V.: A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. IEEE Trans. Pattern Anal. Mach. Intell. 25, 265–270 (2003)
7. Wink, O., Niessen, W., Frangi, A., Verdonck, B., Viergever, M.: 3D MRA coronary axis determination using a minimum cost path approach. Magnetic Resonance in Medicine 47, 1169–1175 (2002)
8. Sethian, J.: A Fast Marching Level Set Method for Monotonically Advancing Fronts. Proc. Nat. Acad. Sci. 93, 1591–1595 (1996)
9. Baltaxe Milwer, M., Flórez Valencia, L., Hernández Hoyos, M., Magnin, I., Orkisz, M.: Fast marching contours for the segmentation of vessel lumen in CTA cross-sections. In: Conf. Proc. IEEE Eng. Med. Biol. Soc., Lyon, France, pp. 791–794. IEEE, Los Alamitos (2007)
10. Hameeteman, K., Zuluaga, M., Joskowicz, L., Freiman, M., van Walsum, T.: Carotid Lumen Segmentation and Stenosis Grading Challenge. In: MICCAI Workshop 3D Segmentation in the Clinic: a Grand Challenge, MIDAS Journal London, UK (2009), http://hdl.handle.net/10380/3128

Cognitive Hierarchical Active Partitions Using Patch Approach

Konrad Jojczyk[1], Michał Pryczek[1], Arkadiusz Tomczyk[1],
Piotr S. Szczepaniak[1], and Piotr Grzelak[2]

[1] Institute of Information Technology, Technical University of Łódź,
ul. Wólczańska 215, 90-924 Łódź, Poland
{michalp,tomczyk,piotr}@ics.p.lodz.pl
[2] Department of Radiology and Diagnostic Imaging
Barlicki University Hospital, Medical University of Łódź
Kopcińskiego 22, 91-153 Łódź, Poland

Abstract. Rapidly developing *Medical Image Understanding* [1] field requires fast and accurate context and semantic oriented object recognition methods. This is crucial due to both description's substantialness requirements and diagnostic responsibility. Synthetic approach to medical image analysis is expected, integrating various kinds of knowledge, to facilitate processing and make the results more meaningful. *Cognitive Hierarchical Active Partitions* is a flexible image analysis tool, facilitating use of semantic and contextual knowledge encoded in patch based linguistic description of a given image. In the paper presented, this technique is evaluated in ventricular system recognition task on example set of brain CT scans.

1 Introduction

Semantic image analysis can be regarded as a process in which sementically meaningful fragments are extracted from an image. Such a fragment is called spatch, which is a blend of spatial and patch to emphasize that it can be used not only in two-dimensional case. The fragments can have different meanings, depending on the analysis level (e.g. blue patch, sky etc.) and domain considered (e.g. bright patch, tumour, etc.). Before the image content is understood, semantic meaning is assigned only to basic elements of the image, i.e. pixels. Only by further analysis supported by additional knowledge can one retrieve complex structures, which then can form other semantically meaningful structures. This is regarded as a bottom-up cognitive process.

The present paper applies *Cognitive Hierarchical Active Partitions* (CHAP), an approach first introduced in [2]. First, in a deterministic way circular patches of approximately the same colour, which are used as a basis for further analysis, are sought. Then, iterative optimisation process, supported by an expert knowledge about the shape, location and colour, retrieves the subset that represents an image fragment which is new from the semantic point of view. The approach

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 35–42, 2010.

presented is innovative in that it takes into account additional knowledge about mutual relations between patches, which places it among systems that apply linguistic knowledge about an image [1,3,4].

2 Linguistic Based Object Recognition

Linguistic-based object recognition is one of the most rapidly developing research domains connected with pattern analysis. Like in structural pattern analysis, an object processed is assigned another object which is to describe it in possible most exact way. In the present paper, the describing objects are circle graphs, generated automatically by using the algorithm presented in [2]. The choice of graph language as a tool of linguistic description is not accidental. Being formal and semantically exact, it is also expressive enough to carry both information about the object's structure and external knowledge about the object.

2.1 Circle Graph

Let $G = (V, E)$ be an undirected graph of linguistic description of an image I. For any vertex $v \in V$ let us define a sphere $K(v)$ in pixel coordinate space of image I, centred at point $O(v)$ and of radius $r(v)$. In the case of two-dimensional images, considered in this paper, each sphere is in fact a circle overlapping some area of I, and, as a result it carries spatial information. Two vertices $w, v \in V$ are adjacent in G if and only if $K(w)$ and $K(v)$ are adjacent (within admissible error margin).

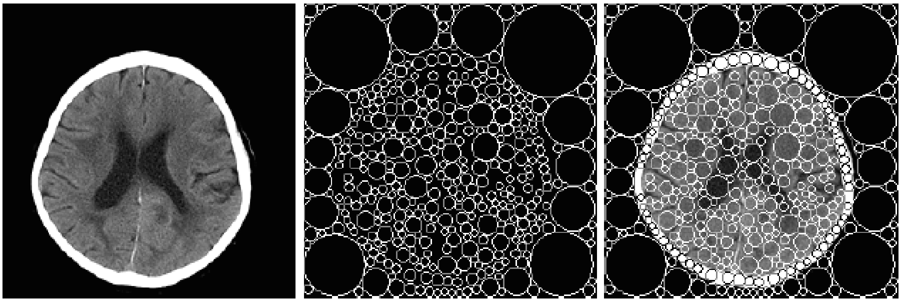


Fig. 1. Description of example image

2.2 Spatio-semantic Information

A fundamental characteristic of a linguistic description is its information content. In the case of a circle graph, spheres defined for each node, contain information about subareas of I which are cohesive colour-wise. Edges carry information about neighbourhood of such subareas and the possibility of their composition into larger areas which may constitute potential candidates in the recognition process. Above all, however, the edges carry information about the shape of an area, supplementing superficial information contained in a set of vertices and associated circles.

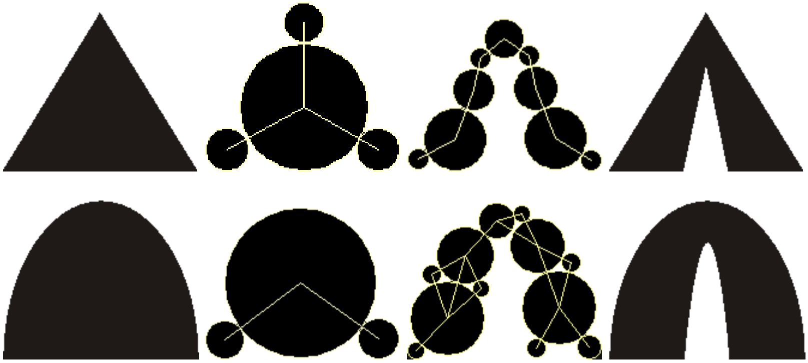


Fig. 2. Verticle and edge carry spatial and shape information

Following the nomenclature introduced in the previous section, let G be a linguistic description of an image I. Assuming that the description is a collection of descriptions of objects distinguishable on I at a given precision level, we can set a task of searching subgraphs G_c of graph G that would fulfil the criteria taken. In the present paper, the criterion taken is described by a prototype graph G_p. Supported by analogous functions K, O and r, the prototype graph carries shape information. Note that the circles are described in abstract coordinate space, which enables one to „put" the prototype in any point of the image in order to compare it with any candidate graph G_c.

3 Linguistic Similarity Measure

The core of the method presented is an algorithm which evaluates the similarity degree between two graph linguistic descriptions: of the candidate (G_c, K_c, O_c, r_c) and of the prototype (G_p, K_p, O_p, r_p).

3.1 Graph Transformations

The algorithm of evaluation presented in 4.1 is based on the search for homeomorphism between graphs G_c and G_p. The algorithm performs matching operations which aim at possibly most exact reflection of G_p structure with the components of G_c. Every matching operation is assigned an edit cost, which helps evaluate the similarity of those shapes which are not homeomorphic according to the definition.

Algorithm of graph description analysis was presented in detail in [2]. In search of homeomorphism of graphs [5,6], two basic edit operations are performed, namely path substitution and path deletion.

3.2 Matching Cost

Both operations, which are multi-stage, use path segment approximation, node substitution, node deletion and path segment deletion. Because every such operation results in graph's modification, it is assigned a cost, estimated in the context of graph descriptions. The final measure of similarity of two graph descriptions (G_p, K_c, O_c, r_c) and (G_c, K_c, O_c, r_c) is the weighted sum of the costs of edit operations performed during the matching of G_p and G_c. Note that there are many matching possibilities of the graphs that differ in cost. The measure is minimum matching.

$$cost(G_c, G_p) = \sum_i w_i c_i \qquad (1)$$

where c_i denotes the cost of a subsequent edit operation, whereas w_i is the weight coefficient of the edit operation, which enables the algorithm to be matched with a class of descriptions considered.

4 Evaluation

During the evaluation process, the algorithm is to locate one of the ventricles of the brain on a computed tomography picture. The analysis of brain ventricles is crucial from a diagnostic point of view, since any deformation of the ventricular system, including compression or expansion of the ventricles or their asymmetry is a factor indicating a disease or pathological changes in the brain.

In the study of the left lateral ventricle, a series of prototypes have been prepared , the following of which have been classified for further research:

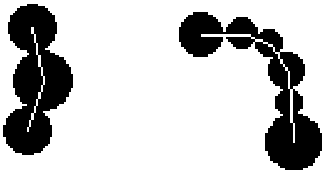


Fig. 3. Knowledge driven prototypes

The search through subgraph's spaces is performed by simulated annealing algorithm with the movement generator which performs random graph change, on the condition that the resultant graph must be connected.

Procedure. `GraphSimilaritySearch`(G_p,G_c) $\in \mathbb{R}$

```
/*Initial match is 1-1 matching between one node from $G_p$ and one
  from $G_c$; while computing costs prototype is considered to be
  placed on the candidate so that centers of initially matched nodes
  overlap. */
$M$ - set of all initial 1-1 matches
while true do
    /*remove worst partial matches; heuristically reduces memory
      consumption ant time complexity */
    trim($M$)
    $best \longleftarrow$ match from $M$ having smalles cost
    remove( $best$ from $M$)
    /*Extending paths has been described in [5]. In short, searching
      homeomorphism between $G_p$ and $G_c$ requires searching for
      mappings between features of both graphs (here edges). */
    if best has extending paths then
        foreach not previously matched $path_p$ extending best do
            /*it is possible to leave $path_p$ unmatched, or to check all
              possibly matching paths in $G_c$ */
            $best_{np} \longleftarrow best$ with $path_p$ matched to nothing
            cost($best_{np}$) $\longleftarrow$ cost ($best$) + penalty for unmatched $path_p$
            add($best_{np}$ to $M$)
            foreach $path_c$ in $G_c$ do
                $best_+ \longleftarrow best$ extended with $path_p \rightleftharpoons path_c$ match
                cost($best_+$) $\longleftarrow$ cost($best$) + ($path_p, path_c$) match cost
                add($best_+$ to $M$)
    else
        $cost_{old} \longleftarrow$ cost($best$)
        cost($best$) += unmatched prototype edges penalty
        cost($best$) += unmatched candidate edges penalty
        if $cost_{old}$ = cost(best) then
            /*cost($best$) didn't change, so it smaller than any other
              processed matching, finishing */
            return cost($best$)
        else
            add($best$ to $M$)
```

4.1 Heuristic Graph Comparison Algorithm

Optimal matching is performed by the algorithm presented above. It is based on heuristic iterative expansion strategy of best possible graph matching.

4.2 Expert Knowledge Consideration

The information about shape similarity is not enough for calculating the correct solution energy. There may be more sets of circles, which correspond to the prototype, but are not necessarily located in the ventricular system.

On the basis of semantic knowledge of an expert, additional optimised energy components can be defined:

- *blackness* - $B(G_c)$ - its value is an average of the values of the pixel colours situated in central points of the circles divided by 255. If the result is lower than the threshold (empirically estimated as 0.2), then the final blackness value is 0. Note that the image is an 8-bit greyscale image, in which black corresponds to 0.
- *centrality* - $C(G_c)$ - its value is an average distance between circles' centres and the centre of the image divided by half of the image's diagonal. On computer tomography images, the head is usually situated just underneath the centre. For this reason, the central point is defined by half width and 0.55 height of the image (counting from the top).

The energy function optimised during subdescription space search is the sum of matching cost [1] and the values of blackness and centrality of the candidate.

$$E^p(G_c) = cost(G_p, G_c) + B(G_c) + C(G_c) \tag{2}$$

Fig. 4 presents exemplary search errors resulting from the omission of energy component.

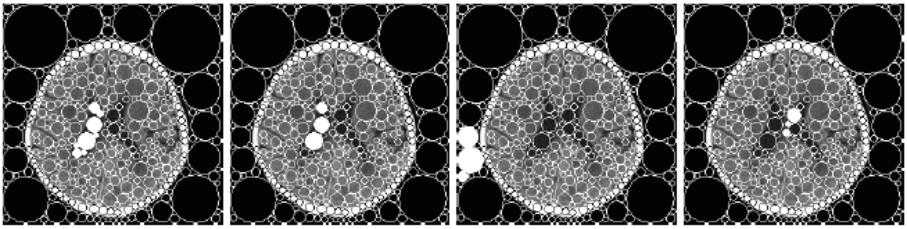

(a) Gold standard (b) $cost + B + C$ (c) Missing C (d) missing B

Fig. 4. Example recognition result

4.3 Research Methodology and Effectiveness Measures

The research is based on a selected set of pictures which clearly show ventricular system of the brain. By means of a standard circle algorithm [2] it is possible to automatically retrieve image circles, then manually determine subgraphs describing the left lateral ventricle. Graph descriptions determined in this way constitute the recognition standard.

The efficiency evaluation of the recognition algorithm was performed by the standard precision and recall, which have been combined with measure F1, averaging the results for 10 separate recognition processes.

4.4 Results

In the first stage of the research, general correctness of the algorithm was verified. The prototype used here was the description which exactly corresponded to the set of circles we were searching for. The subgraph was found correctly each time. In the second stage, the prototypes were expert graphs based on medical knowledge about the shape of left ventricle. Both are presented in Fig. 3. Each image was assigned a prototype which approximately corresponded to the shape presented in the picture. The results are depicted in Table 1.

Table 1. Results of evaluation on selected images

Image	Measure	Precission	Recall	F1	Image	Measure	Precission	Recall	F1
norm 1	mutualinf	0.8	0.46	0.58	path. 1	mutualinf	0.7	0.41	0.51
	purity	1	0.58	0.73		purity	0.6	0.33	0.43
	fmeasure	0.9	0.52	0.73		fmeasure	0.4	0.25	0.30
	entrophy	0.8	0.52	0.73		entrophy	0.5	0.33	0.40
norm 2	mutualinf	0.49	0.61	0.54	path. 2	mutualinf	0.9	0.43	0.58
	purity	0.45	0.55	0.49		purity	0.7	0.33	0.45
	fmeasure	0.49	0.61	0.54		fmeasure	0.7	0.35	0.46
	entrophy	0.5	0.61	0.55		entrophy	0.9	0.43	0.58

Basing on the conducted experiments it can be observed that the method presented correctly localises chambers of ventricular system. Yet, the exact segmentation is still unsatisfactory if precie area or shape recognition are expected. This can be, however, considered as the first step in image processing, where the skeleton of an object described by the graph is later improved. Moreover, it can be noticed, that not all of the results presented in Table 1 are satisfactory. It can be explained by the wrong choice of the energy parameters. Depending on the size of the circles in graph descriptions unmatching penalties used by the *GraphSimilaritySearch* procedure must be properly adjusted in order not to marginalise circles with small diameters.

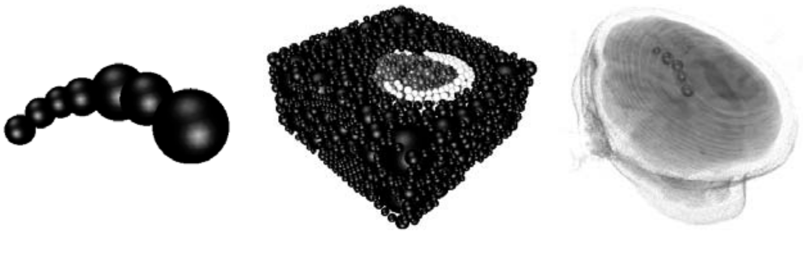


Fig. 5. Example prototype, description and results in 3D space

5 Conclusions and Further Work

The present paper applies graph linguistic description of images to object retrieval and recognition in the image. The method presented is characterised by high efficiency not only when exact prototype description is provided. It was designed in such a way as to retrieve even shapes described in an inexact or superficial way. However, in the latter case, efficiency depends to a large extent on the object's characteristics and the context. It is worth mentioning that the method presented can be also easily applied in multidimensional space. Fig. 5 visualises the same description method applied in three-dimensional space.

Acknowledgement

This work has been supported by the Ministry of Science and Higher Education, Republic of Poland, under project number N 519007 32/0978; decision no. 0978/T02/2007/32.

Authors would like to express their gratitude to Mr Paweł Karpiński, for inspiring concepts regarding linguistic description generation.

References

1. Tadeusiewicz, R., Ogiela, M.: Medical Image Understanding Technology. Studies in Fuzziness and Soft Computing, vol. 156. Springer, Heidelberg (2004)
2. Tomczyk, A., Jojczyk, K., Pryczek, M., Szczepaniak, P.S.: Spatch based active partition with linguistically formulated energy. Journal of Applied Computer Science 18(1) (2010) (accepted for publication)
3. Gonzalez, R., Woods, R.: Digital Image Processing. Prentice-Hall Inc., New Jersey (2002)
4. Sonka, M., Hlavec, V., Boyle, R.: Image Processing, Analysis and Machine Vision. Chapman and Hall, Cambridge (1994)
5. Rocha, J., Pavlidis, T.: A shape analysis model with applications to a character recognition system. IEEE Trans. Pattern Anal. Mach. Intell. 16(4), 393–404 (1994)
6. Rocha, J., Pavlidis, T.: Character recognition without segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 17, 903–909 (1995)

Ontological Models as Tools for Image Content Understanding

Juliusz L. Kulikowski

Institute of Biocybernetics and Biomedical Engineering PAS
4 Ks. Trojdena str., 02-109 Warsaw, Poland
juliusz.kulikowski@ibib.waw.pl

Abstract. An approach to computer-assisted image content understanding is presented. It is based on using domain ontology as a source of knowledge concerning the external world whose objects are visualized by the image. The ontology is presented in the form of ontological models constructed on the basis of widely-defined relations. Image content is recognized through solution of relational equations. The main steps of computer-assisted image interpretation and their content understanding are illustrated by examples.

Keywords: Image processing, image content understanding, domain ontology, relations, relational equations.

1 Introduction

Image can be considered as multidimensional (2D, (2+1)D, 3D, (3+1)D, etc.) signal characterised by a multi-level organisation. In a computer-aided image processing context the 1/ physical (spectral), 2/ syntactic (statistical, morphological, structural), 3/ semantic, 4/ pragmatic and 5/ systemic image organization levels can be distinguished. For the first decades of image processing methods development attention on the physical and syntactic image properties was focused. However, in many application areas (science, technology, medicine, arts etc.) where important role by visual information is played a computer support to image content recognition, image practical usefulness assessment, detection of their relationships to other documents stored in the systems, etc. is also highly desired. That is why the interest to the semantic, pragmatic and systemic aspects of computer-aided image processing for the last decades was growing up [10]. Solution of any of the above-mentioned problems needs a deal of information about the application area to be taken into consideration. There are many ways of explaining what image understanding means. E.g., in a large context, J. Hawkins and S. Blakeslee [7] explained *understanding a situation* as an *ability of a living individual to preview the life-important effects of the situation.* In this sense image understanding includes recognition of semantic, pragmatic as well as systemic aspects of the image under consideration. R. Penrose [18] additionally stresses the role of consciousness in understanding a situation. Below, following the J. Hawkins and S. Blakeslee concept (and without going into philosophical

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 43–58, 2010.

deliberations whether it is or is not ,,conscious" (the more that *consciousness* also has not been by R. Penrose exactly defined), we define *image content understanding as ability: 1st to interpret image in the terms of an assumed model of external world reflected in the image, and 2nd on the basis of image analysis, to reply queries concerning image itself as well as reflected by it objects, their properties, widely defined relationships among the objects, and their relationships with other, directly non-observed objects.* In such sense, *image understanding* becomes a synonym of *rational image interpretation.* Effective image understanding is strongly dependent on a *primary knowledge* about a given application domain (a selected part of real or abstract world under consideration) and about the image acquisition or creation circumstances. The above-mentioned types of primary knowledge may take various forms. In the simplest case they may take the form of less or more *formalized models.* For example, non-formalized models which in ancient astronomy were used to explain configurations of stars visible on a night sky had the forms of animals (Lion, Taurus, Great Bear, Small Bear, etc.) or of mythological figures (Orion, Archer, Gemini, etc.). This sort of external world models helping sky image interpretation made a preview of the position of given stars constellations possible. However, it was unusual for the positions of moving planets, comets, etc. previewing. A well-known sort of formalized external world models used in pattern recognition problems based on Bayesian rule has the form of a priori probability distributions of the patterns. In this case image interpretation is limited to indication of the similarity class of objects the presently recognized one probably belongs to. Let us suppose that for each recognized pattern (similarity class) links to some pattern properties, practical aspects, etc. are given. Then, after the pattern recognition act all attached data can be presented to the user. Consequently, in principle, answering many questions on the basis of those data becomes possible. Does it mean that the recognizing system ,,understood" the content of the image? This example shows that despite the formerly given explanation of what image content understanding means the borders of meaning still remain fuzzy. On one hand, the correctness of understanding can be evaluated only within a given, more general concept of the external world (the previously assumed models of the application area); on the other one, the depth of understanding depends on the class of questions we expect to be answered on the basis of image analysis. From a user's point of view it is not so much important that the image content interpreting system acts ,,consciously", it is rather important that it correctly replies to the queries concerning the image under consideration and presented by it objects and their environment. The replies should be based on the application domain knowledge. The more deep, correct, extended and adequate to the user?s expectations the domain knowledge is, the better computer-assisted image content understanding is possible. The domain knowledge is usually provided to the system by experts in the given application area. Therefore, a computer-based image content understanding system may take the form of a specialized expert system shown in Fig. 1.

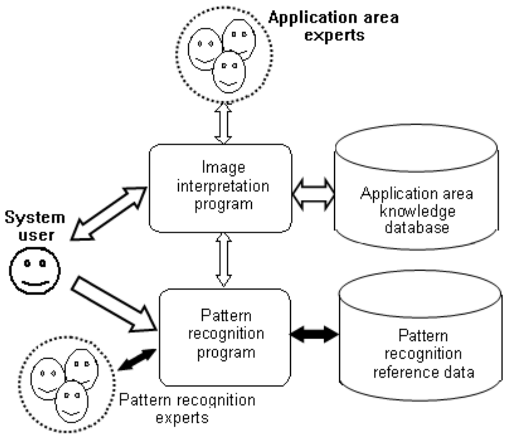

Application area experts
Image interpretation program
Application area knowledge database
System user
Pattern recognition program
Pattern recognition reference data
Pattern recognition experts

Fig. 1. General scheme of an image's content understanding system

In this scheme, besides the user, two groups of experts have been indicated. The first one provides the system with reference data used to pattern recognition. The second one provides to image understanding necessary application area knowledge. The user does not need to get direct contact with any of those groups. He communicates with the system by two ways: one concerning image introduction and recognition process control and another one concerning image interpretation, i.e. putting questions about the objects visible in the images rather than the images themselves and evaluating the system replies. The communication tracts concerning image interpretation procedures have been marked in the scheme by enlarged arrows.

Our interest will be focused below on a concept of the image interpretation program and application area knowledge database. In this context two main problems arise: 1^{st}, what should be the form of application area knowledge representation in the system, and 2^{nd}, how to use this knowledge for answering the user's queries concerning the observed objects and their relationships with other application area objects. Below, a short view on a concept of the above-formulated problems solution will be given. It is based on a general concept of *ontology* as a form of domain knowledge representation which in the last decades became popular in various application areas [1, 2, 5, 6]. The concept of using ontologies in pattern recognition was formulated in [12] and developed in [13, 14, 15, 16] where a particular attention to the relations as mathematical tools for ontological models construction was paid. However, it should be remarked that for the last years many other approaches to the image content understanding problem have been proposed in the literature, e.g. [3, 4, 8, 17, 19]. It seems that until a significant number of non-trivial experimental results of computer-aided image content understanding has been gathered no strong preferring of some approaches to the problem with respect to the other ones will be sufficiently plausible.

The paper is organized as follows. In Sec 2. the idea of ontological models as tools of application area knowledge representation is shortly reminded and a concept of their general structure is proposed. In Sec. 3 the way of ontological models representation using the formalism of relations and hyper-relations is described. Sec. 4 contains an example of USG cardiac images interpretation method using elements of cardiologic ontology and corresponding ontological models. Conclusions are gathered in Sec. 5.

2 Ontological Models

The below-presented approach to image content understanding is based on several basic assumptions.

2.1 Domain Ontology

The application domain knowledge needed to image content understanding can be presented in the form of *domain ontology* O specifying a taxonomy T of basic *concepts* and a family of *relationships* between the concepts representing various aspects of the given application area of interest. By the concepts we mean any real or abstract individual objects or their ensembles (classes of similarity). The qualitative attributes of objects and numerical parameters characterising the objects can be considered as concepts as well.

For pattern recognition and image analysis purposes it is appropriate to represent the domain ontology O in a formalized way as an ensemble of *ontological models (OM)*:

$$O = \{OM_1, OM_2, \ldots, OM_I; Q, A\} \tag{1}$$

where Q denotes a subset of super-relations between the OMs describing a general structure of the ontology and A is a set of axioms concerning the models (e.g. extended algebra of relations or hyper-relations, probabilistic assumptions, etc.). Ontological model can formally be defined as a quadruple:

$$OM_i = [C_i R_i, Top_i, A_i], \quad i = 1, 2, ..., I, \tag{2}$$

such that: C_i is a non-empty subset of concepts, R_i is a super-relation describing the elements of C_i containing, in particular, a multi-aspect taxonomy Ξ_i of the elements of C_i; $Top_i \in C_i$ being its highest element; A_i is a subset of model assertions. Using the notions of ontological models a domain ontology O can be constructed in a bottom-up process by construction and completing ontological models adequate to the description of the given application domain. In particular, a multi-aspect taxonomy can be defined as a set of taxonomic trees:

$$\Xi_i = \{T_i^{(1)}, T_i^{(2)}, \ldots, T_i^{(m)}, Top_i\} \tag{3}$$

having a common root Top_i and classifying the elements of C_i from different points of view.

The set of concepts of a given ontology:

$$C = \cup_i C_i \tag{4}$$

can also be presented by a sum of three (not obviously non-overlapping) classes:

$$C = C^{(im)} \cup C^{(ob)} \cup C^{(ad)} \tag{5}$$

where $C^{(im)}$ denotes concepts directly concerning the analyzed image, $C^{(ob)}$ - the concepts concerning objects belonging to the application area visualized by the image, and $C^{(ad)}$ - other concepts concerning the application domain, indirectly related to the analyzed objects. On the basis of so-defined concepts three types of ontological models as shown in Fig. 2 can be constructed: 1st those image describing in statistical, geometrical, spectral, morphological or structural terms, 2nd those assigning application area concepts (terms) to the identified image details, and 3rd those interpreting the recognized objects' properties and the relationships between them and other, non-observed directly objects belonging to the application area and constituting a large image analysis context.

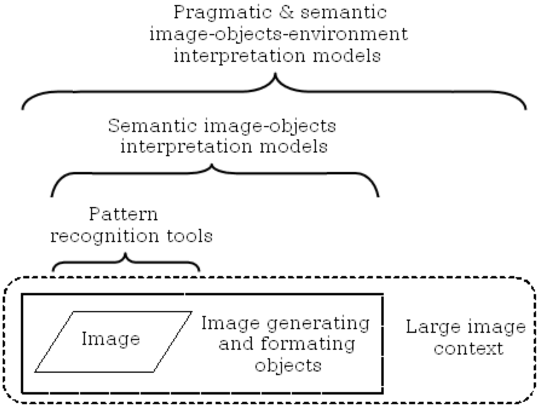


Fig. 2. Hierarchy of image interpretation models

The idea of a domain ontology construction can be illustrated by the following example.

Example 1

Let us assume that practical problems consist in computer-aided interpretation of medical cardiac images. The application area thus can be defined as *Cardiology* and the corresponding taxonomy of concepts can be displayed in a taxonomic tree whose upper layers may have the following form:

Cardiology:

- *Heart,*
  - ○ *Heart anatomy,*
  - ○ *Cardiac physiology,*
  - ○ *Cardiac pathology*
- *Heart treatment,*
  - ○ *Prevention methods,*
  - ○ *Diagnostic methods,*
  - ○ *Pharmacological methods*
  - ○ *Surgical methods,*
  - ○ *Rehabilitation*
- *Cardiologic drugs and instruments,*
- *Cardiologic medical care,*
- *Cardiac patients,*

etc.

In cardiac image processing various concepts may be of particular interest. One can reach them following selected paths on the taxonomic tree, like e.g.:

Cardiology - Heart -

- *Heart anatomy - Cardiac chambers - Left cardiac ventricle,*
- *Pathology - Cardiac failure - Ventricle's contractility disorder - Akinesis,*

etc. In similar way, we can draw the paths:

Cardiology - Heart treatment -

- *Cardiac diagnostic methods - Cardiac imaging analysis - USG modality -*
- *Kinetic contours description - Kinetic parameters;*

Cardiology - Cardiologic medical care - Cardiac examinations -

- *Visual diagnosis - Visual examination file - Radiological examination file,*
- *Electrocardiography - ECG. examination file - Electrocardiogram,*

etc.

In practice, construction of a complete taxonomy of concepts of an application domain is a laborious task needing collaboration of a large number of specialists. However, for narrow classes of image interpretation tasks solution sub-taxonomies based on existing repertories of practically used terms and/or concepts can be used. On the other hand, a clearly defined hierarchy of concepts concerning images and related to them application areas makes construction of ontological models adequate to image content understanding much easier.

2.2 Relations

The relations occurring in ontological models are considered in a wide sense including extended, conditional, parametric relations and super-relations (relations described on lower-level relations considered as sets) [9], as well as hyper-relations (sums of [11]). From another point of view, they may be of deterministic, probabilistic or indeterministic (e.g., on multi-valued, fuzzy or relative logic based) type. Due to this, relations become a large and flexible tool for a wide class of real situations description.

Taking into account the above-mentioned three classes of concepts all relations in the ontological models can be divided into the following groups:

a) Relations described on the concepts of $C^(im)$ only, characterizing structural properties of images;
b) Relations described on the concepts of $C^{(im)} \cup C^{(ob)}$, characterizing the relationships between images and visualized objects, making the recognition of image content possible;
c) Relations described on the concepts of $C^{(ob)} \cup C^{(ad)}$, characterizing the inter-objects relationships within the application domain and making possible an extended semantic, pragmatic or systemic image aspects analysis.

The above-given assumptions make construction of a large class of ontological models possible. However, for effective image content understanding system construction the domain ontology should be flexible and extendable. Moreover, despite the fact that the ontology-based approach to image understanding is mainly declarative, fitting it with some elements of a procedural approach seems reasonable. This can be reached due to several possible ways of relations description:

- ○ by listing instances (elements of the corresponding Cartesian product, n-tuples) satisfying the relation;
- ○ by algebraic compositions of other, formerly described relations;
- ○ by using testing procedures to select instances satisfying the relation.

Choosing an adequate structure of application domain ontology is thus a crucial point in ontology-based image content understanding. Established domain ontology concepts should put a bridge between the world description used by computer specialists and this of the application domain experts.

Example 2
Let us assume that the following concepts have been established: *Patient identifier (Pid), Examination date (Exd), Examination file (Exf).* They are specified as the following sets:

$Pid = \{AK, KL, NP, WZ\}$ - (symbols stand here for shortened names of patients),
$Exd = \{01 - 01, ..., 31 - 12\}$ - (all days round the year),

$Exrf = \{a/001, ..., a/999\}$ - (identifying symbols of radiological examination files)
$Exef = \{b/001, ..., b/999\}$ - (identifying symbols of ecg examination files).

On the basis of the above-given sets two relations describing radiological ($R1$) and electrocardiographic ($R2$) examinations by listing their instances are described:

$$\begin{aligned} Pid \times Exd \times Exrf \supseteq R1 \equiv & \{(AK, 12-02, a/127), (NP., 14-02, a/131), \\ & (WZ, 16-02, a/137)\}, \\ Pid \times Exd \times Exef \supseteq R2 \equiv & \{(NP, 06-02, b/328), (AK, 12-02, b/329), \\ & (AK, 12-02, b/331), (KL, 12.02, b/343)\}. \end{aligned}$$

Then, on the basis of an extended algebra of relations [4] there can be defined:
• An extended sum of relations:

$$\begin{aligned} Pid \times Exd \times Exrf \times Exef \supseteq R1 \cup R2 \equiv & \{(AK, 12-02, a/127, *), \\ & (NP., 14-02, a/131, *), \\ & (WZ, 16-02, a/137, *), \\ & (NP, 06-02, *, b/328), \\ & (AK, 12-02, *, b/329), \\ & (AK, 12-02, *, b/331), \\ & (KL, 12.02, *, b/343)\}. \end{aligned}$$

Stars * are used for denoting any element of the corresponding set.
• An extended intersection of relations:

$$Pid \times Exd \times Exrf \times Exef \supseteq R1 \cap R2 \equiv \{(AK, 12-02, a/127, b/329)\},$$

• Extended differences of relations:

$$\begin{aligned} Pid \times Exd \times Exrf \times Exef \supseteq R1/R2 \equiv & \{(NP, 14-02, a/131, *), \\ & (WZ, 16-02, a/137, *)\}. \\ Pid \times Exd \times Exrf \times Exef \supseteq R2/R1 \equiv & \{(NP, 06-02, *, b/328), \\ & (KL, 12.02*, b/343)\}. \end{aligned}$$

• Extended symmetrical difference of relations:

$$\begin{aligned} Pid \times Ext \times Exd \times Exrf \supseteq R1 \div R2 \equiv & \{(NP, 14-02, a/131, *), \\ & (WZ, 16-02, a/137, *), \\ & (NP, 06-02, *, b/328), \\ & (KL, 12.02, *, b/343)\}. \end{aligned}$$

• Extended conditional relations:

$$\begin{aligned} Pid \times Ext \times Exd \times Exrf \supseteq R1|R2 \equiv & \{(AK, 12-02, a/127), *)\}, \\ Pid \times Ext \times Exd \times Exrf \supseteq R2|R1 \equiv & \{(AK, 12-02, *, b/329), \\ & (AK, 12-02, *, b/331)\}. \end{aligned}$$

2.3 Super-Relations

For the ontological models construction and description purposes the names of relations as well as of their arguments are given as symbolic denotations of concepts provided by ontological taxonomic trees. For example, above, in Example 2 there have been defined relations which can be denoted as $R1(Pid, Exd, Exrf)$ and $R2(Pid, Exd, Exef)$. However, some of relations' arguments can be defined as relations themselves. In the above-given case the argument Exd is, in fact, a relation $Exd(Day, Month)$ where Day is a concept defined as a set of integers $Day = \{1, 2, 3, \ldots, 31\}$ and, similarly, $Month = \{1, 2, \ldots, 12\}$. Therefore, the former relations could also be written as $R1[Pid, Exd(Day, Month), Exrf]$ and, respectively, $R2[Pid, Exd(Day, Month), Exef]$. The last form of relation directly indicating some of its arguments as lower-level relations is called a super-relation.

Example 3
Let us assume that it is defined a relation $Patient(Pn, Pid, Pbd, Pad)$ where Pn denotes patient's name, Pid is patient's identifier (look Example 2), Pbd is patient's birth date, and Pad is patient's address. Pbd and Pad can be extended into lower-level relations. It will be also taken into consideration the formerly-defined relation $R1 \equiv R1(Pid, Exd, Exrf)$. Then, a conditional relation $Patient|R1$ will consist of all instances of the relation Patient such that there exist in $R1$ instances containing the same patient's identifier (Pid), while an intersection $Patient \cap R1$ is a relation depending on the arguments ($Pn, Pid, Pbd, Pad, Exd, Exrf$).

2.4 Hyper-relations

The number and order of arguments of classical relations are strongly fixed. In ontological models construction this constraint is rather undesired. The concept of hyper-relations tries to overcome this inconvenience [18]. Let F be a family of $|F|$ non-empty sets (below, it will be assumed that $|F|$ is a finite natural number). We take into consideration a family $G = 2^F$ of all sub-families of F and a family B_G of all linearly ordered permutations of the elements of G. Each element of B_G is thus a linearly ordered sub-family of the sets belonging to F and as such, it may generate some relations. We denote by R_F the family of all relations described on any elements of B_G. Then any set of instances of any relations belonging to R_F is called a *hyper-relation* (a h-relation) described on the family F of sets. The instances of a h-relation are thus some sequences of the elements of sets belonging to linearly ordered sub-families of F, the admissible lengths of sequences varying from 0 to $|F|$.

Example 4
Let us take into consideration the above-defined relation $R1(Pid, Exd, Exrf)$. We shall denote by $R1(NN, *, Exrf)$ its sub-relation consisting of all instances

for which $Pid = NN$ is strongly fixed and $Exd = *$ denotes any admissible examination date. Then, we select a *projection* of this sub-relation on the argument $Exrf : Q = R1(NN, Exd, Exrf)|_{Exrf}$ being a set of all radiological examination files acquired at any date from the patient NN. The examination files are given in the form of sequences of radiologic snapshots (frames). The sequences consist of time-ordered sub-sequences of various lengths presenting the given cardiac action in fixed projections. All such sequences are thus a sub-family of the above-mentioned family B_G of linearly ordered sequences of elements drawn without repetitions from a set C of all radiologic snapshots under consideration. Then, a set of all sub-sequences of radiologic snapshots presenting cardiac action in a fixed projection is a h-relation described on the family F of sets C (taken as many times as the maximal length of sequences of frames is).

2.5 Parametric and Fuzzy Relations

If $r(C_1, C_2 \ldots C_k)$ is a relation described on a Cartesian product $C_1 \times C_2 \times \ldots \times C_k$ then formally it is easy to assign a numerical parameter $\mu, \mu \in M$, to its instances and to replace the relation by its extended form $r^*(C_1, C_2, \ldots, C_i,$ $\ldots, C_k, M)$ described on a Cartesian product $C_1 \times C_2 \times \ldots \times C_i \times \ldots C_k \times M$. This form of relation will be called a *parametric relation.* The parameter μ may in general provide any sort of *meta-information* about the relation instances, e.g. their costs, importance, credibility, etc. In principle, they also may represent probability of the instances considered as random events or *membership coefficients* when the relation is considered as a fuzzy in L. Zadeh sense relation. However, in the last cases the values of the parameter assigned to the relation instances should satisfy the normalization constraints: $\Sigma_i \mu_i \equiv 1$ being itself a hyper-relation described on a countable Cartesian product of the sets M. We denote this hyper-relation by $Norm(M)$. Therefore, it is better to consider a probabilistic or a fuzzy relation as *conditional parametric relation* of the form $r^*(C_1, C_2, \ldots, C_i, \ldots, C_k, \mu)|Norm(\mu)$ which is satisfied by such instances of $r^*(C_1, C_2, \ldots, C_i, \ldots, C_k, \mu)$ for which the normalization condition by the parameter values μ is satisfied.

2.6 Relational Equations

For a given relation $r(C_1, C_2, \ldots, C_i, \ldots, C_k)$ described on a Cartesian product $C_1 \times C_2 \times \ldots \times C_i \times \ldots \times C_k$ where $C_1, C_2, \ldots, C_i, \ldots, C_k$ denote some sets an expression:

$$r(a_i, b_j, \ldots, x, \ldots, e_k) = t. \tag{6}$$

where $a_i \in C_1, b_j \in C_2, \ldots, x, \ldots, e_k \in C_k$, $t \in \{$„true", „false"$\}$, can be interpreted as an equation with respect to the variable x. Its solution is any $x_0 \in C_i$ such that $(a_i, b_j, \ldots, x_0, \ldots, e_k)$ is an instance of the relation r.

Relational equations make formulation of questions for a relations-based decision system possible, because (6) is semantically equivalent to the query: „Are there any, and if so, what are the elements x of C_i such that the relation $r(a_i, b_j, \ldots, x, \ldots, e_k)$ takes assigned to it logical value".

Example 5
Let us take into consideration the above defined relations $R1(Pid, Exd, Exrf)$ and $R2(Pid, Exd, Exef)$. A set of relational equations:

$$\begin{cases} R1(x,(12,03),Exrf) = \text{,,}\textit{true}\text{''} \\ R2(x,(12.03),Exef) = \text{,,}\textit{false}\text{''} \end{cases}$$

is semantically equivalent to the query: ,,*Are there any, and if so, what are the identifiers of patients who have passed a radiological examination on 12.03 and have not passed the same day an ECG examination*".

The concept of single-variable relational equation can easily be extended on multi-variable relational equations as well as on sets of single- or multi-variable relational equations. Basic properties of relational equations still need to be investigated. Some remarks concerning the way to solve the sets of relational equations will be given below.

2.7 General Structure of a Domain Ontology

A general structure Q of the domain ontology O can be described by a set of super-relations between the ontological models OMs. Two types of super-relations can be mentioned as the most important ones. The first one concerns expression (3) showing that, unlike the widely used definitions of ontologies [1, 2, 6], construction of a domain ontology needs sometimes using several taxonomic trees to show the hierarchies of concepts based on various approaches. For example, a taxonomy of drugs can consist of particular taxonomies based on chemical components, therapeutic destination, way of pharmacological action, etc.

The second type of super-relations represents semi-ordering of relations caused by inclusions of their sets of arguments. If $R'(G')$ and $R^{''}(G^{''})$ are two relations described, correspondingly, on the sub-families of sets G' and $G^{''}$, then it can be called that:

a) $R'(G')isapredecessorofR^{''}(G^{''})$ if $G' \subset G^{''}$,
b) $R'(G')isasuccessorofR^{''}(G^{''})$ if $G' \supset G^{''}$,
c) $R'(G')andR^{''}(G^{''})arerelated$ if $G' = G^{''}$,
d) $R'(G')andR^{''}(G^{''})arenon-related$ otherwise.

The above-defined situations can be, respectively, denoted by $R'(G') \to R^{''}(G^{''}), R'(G') \leftarrow R^{''}(G^{''}), R'(G') \leftrightarrow R^{''}(G^{''})$, and $R'(G')?R^{''}(G^{''})$. Moreover, the non-related pairs of relations can be divided into two sub-classes:

a. $R'(G'), R^{''}(G^{''})areconnected, R'(G')\_R^{''}(G^{''})$, if they are non-relative but $G' \cap G^{''} \neq \emptyset$ where $\emptyset$ is an empty set;
b. $R'(G'), R^{''}(G^{''})aredisconnected, R'(G') \perp R^{''}(G^{''})$, if they are non-relative and $G' \cap G^{''} = \emptyset$.

According to this,

1. $R'(G')$ is a predecessor of $R^{''}(G^{''})$ if it is: a partial relation of $R^{''}(G^{''})$;

2. an extended sum of relations is a successor of its components as well as of its sub-relations;
3. an intersection of extended relations is their predecessor;
4. any relation is a successor of its conditional relations,

etc.

The domain ontology structure Q is thus given by a semi-ordering of relations given by the assumptions a) - d) and, in addition, by a binary super-relation of connection given by the assumption a). Both super-relations can be presented in the form of a graph H whose nodes are assigned to the relations and selected pairs of nodes are linked by three types of edges: arcs → between the predecessors-successors, double arcs ↔ between the related nodes, and undirected lines - between connected nodes. We call H a *structural graph of the domain ontology.*

Example 6
Let us take into consideration the following family of relations:

Patient(Pn, Pid, Pbd, Pad), Pbd(Day, Month, Year), Exd(Day, Month, Year), Pad(Town, Postal-code, #House, #Flat), R1(Pid, Exd, Exrf), R2(Pid, Exd, Exef), Exrf(Pid, Rld), Exef(Pid, Ecgd), Rld(Imf), Ecgd(Ecgf), Cardiag(Pid, Cardis, Rld, Ecgd)

where the meaning of the relations $Patient, R1, R2, Pbd, Exd, Pad, Exrf$ and $Exef$ has been explained in the former examples while the meaning of other relations or sub-relations is given below:

Imf - radiological image file, Ecgf - ECG file, Rld - radiological examination data, Ecgd - electrocardiographic examination data, Cardiag - cardiac diagnosis, Cardis - cardiac disease.

The structure of this part of a domain ontology is shown below in Fig. 3. Existence of a → type or a _ type connection between nodes indicates that the corresponding relations provide information which may be combined through a set of relational equations. For example, a pair $R1\_Cardiag$ having common component Pid and $Exrf$, by fixing the parameters $Pid = NN$ and $Exd = (DD, MM, YY)$ makes it possible to formulate a relational equation:

$$Cardiag(NN, x, Exrf, Exef)|R1(NN, (DD, MM, YY), Exrf) = \text{,,}\textit{true}\text{''}$$

which can be interpreted as follows: *,,What is the disease x of the patient NN diagnosed on the basis of radiological (Exrf) and electrocardiographic (Exef) examination data assuming that (R1) the radiological examination Exrf has been performed on (DD, MM, YY)".* In this case the relational equation realizes a process of extended interpretation of the results of basic image processing represented by a relation $Rld(Imf)$ assigning a set of diagnostic parameters to the original radiological image Imf.

A ↔ type connection in the structure of ontology is useless from the above-presented point of view as connecting relations of similar structures but totally disjoint in semantic sense (like: *Birth date and Examination date*). A sub-graph obtained by removing from the structural graph H all ↔ type edges represents

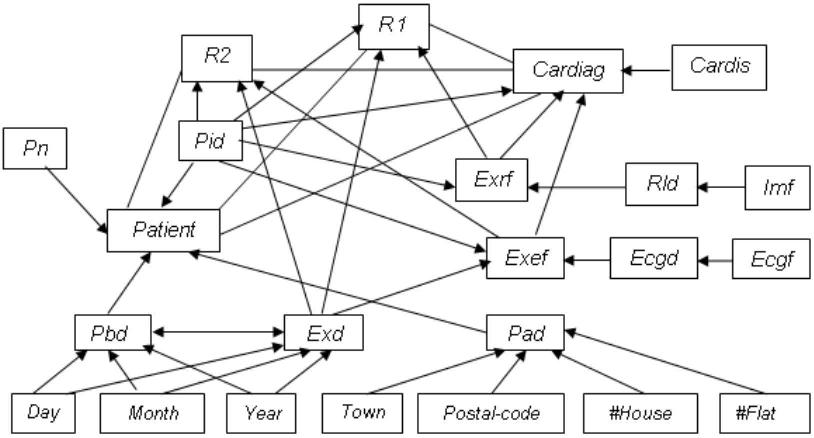


Fig. 3. Relationships between the components of a part of domain ontology

the structure of ontological concepts and relations distinguished within the given application area. We denote the graph by SF and call it a structural semantic field of the given ontology.

3 Image Interpretation

Till image understanding is considered as ability to give rational replies to queries concerning image itself as well as its relationships with other application domain objects, phenomena or processes, the structural semantic field SF determines the borders of the area of possible image interpretations that can be based on the given domain ontology.

The sub-family $C^{(im)}$ of sets mentioned in Sec. 2.1 is a basis of relations describing the properties of analyzed images on their spectral, morphological, statistical, geometrical and/or structural levels. Any relational equations formulated on the basis of this sub-class of relations can be used to reply queries formulated using the concepts concerning formal aspects of the images.

Extension of $C^{(im)}$ on $C^{(ob)}$ leads to the possibility of assigning semantic values to the geometrical, morphological or structural details recognized on the basic image-level. In this case talking about recognized *veins, cardiac chamber contours, moving cars, human faces*, etc. as well as about artifacts like shadows, occlusions, image defects, etc. becomes possible. Relational equations formulated on the basis of the $C^{(im)} \cup C^{(ob)}$-type relations make possible replying the queries like: *,,Is the white spot (in a mammogram) a calcification?"*, *,,Are there any cancerous cells in the given cytological specimen?"*, *,,Which objects (in a series of street traffic photo-shots) are moving cars?"*, etc. In semantic image interpretation associations between application area concepts and pattern recognition concepts are realized through the domain ontology concepts represented by common relation arguments. This can be illustrated by the following example.

Example 7
Let us assume that series of left cardiac ventricle contours are analyzed. Image analysis provides some parameters W based on spectral coefficients S describing the ventricle shape changes along several heart evolution cycles; the parameters W describe in particular the irregularity of the process (which only at a first glance is periodical). The shape changes can thus be formally described by a relation $SHC(W, S)$ assigning *contour changes irregularity parameters* W to the spectral process coefficients S. On the other hand, the domain ontology contains a concept of *Heart contraction (HC)* which is specified by a relation $HC(HCT, W)$ where HCT denotes the concept of *Heart contraction type* (e.g. *Fairly regular, Regular, Irregular, Interrupted, Wasting*). Extended intersection of the relations $HC(HCT, W) \cap SHC(W, S)$ links the corresponding medical and pattern recognition concepts, the concept W playing an associating function. When the relations are given various queries on their basis can be replied, e.g.:

$$\begin{cases} HC(x, y) = \text{,,}true\text{''}, \\ SHC(y, s) = \text{,,}true\text{''} \end{cases}$$

which for given values $S = s$ are interpreted as: *What heart contraction type (x) corresponds to the contour described by the given spectral process coefficient s?"* ;

$$HC(regular, x) = \text{,,}true\text{''}$$

interpreted as: *What contour changes irregularity parameters (x) justify the conclusion that the heart contraction type is regular?"*;

$$SHC(y, x)|HC(irregular, y) = \text{,,}true\text{''}$$

interpreted as: *What are the spectral contour coefficients (x) and the corresponding contour changes irregularity parameters (y) assuming that they led to the conclusion that the heart contraction type was irregular?"* , etc.

For a given image the structural semantic field SF becomes limited to its compact part containing the image (as argument of lower-level relations) and, possibly, some other relations whose instances or selected arguments represent auxiliary knowledge taken into account in image interpretation. So-modified and limited part of SF will be called a *data-oriented structural semantic field* and will be denoted by $d{-}SF$. The $d{-}SF$ determines the area of possible interpretations of the given image, i.e. of formulation queries concerning the image, representing them in the form of relational equations and of reaching replies to the queries in the form of solution of the equations. Of course, the scope of image interpretation, i.e. the variety of admissible queries and replies depends on the scope of the domain ontology, the correctness of replies depends on this of ontological models (relations), their pertinence depends on the pertinence and accuracy of the data entered in the instances of relations. Computer-assisted image interpretation is evidently different than the human one. Computer action is not conscious (whatever does it mean). Moreover, till now, computers cannot create going beyond the domain ontology hypotheses about the image content, while humans can do it and due to this - sometimes go to unexpected and valuable conclusions.

4 Conclusions

Image content understanding is the next step in computer-assisted image processing methods. There are many possible approaches to the problem what ,,image understanding" means and how could it be realized. In any case it needs taking into consideration a large context of image arising and of its relationships to an real or abstract external world. In this paper an approach based on domain ontology concepts and on widely-defined relations used to ontological models construction is presented. This approach leads to a general way to interpret images, their contents and to recognize their relationships with external world through solution of adequately chosen relational equations. However, this general way leads also to many new problems: how to construct ontological models, how to effectively solve relational equations, etc. Basic problems of non-contradiction and existence of solutions of relational equations need also to be investigated.

References

1. Abdoullaev, A.: Reality, Universal Ontology, and Knowledge Systems. In: Toward the Intelligent World. IGI Publishing, Hershey (2008)
2. Bodenreider, O., Burgun, A.: Biomedical ontologies. In: Chen, H., et al. (eds.) Medical Informatics. Knowledge Management and Data Mining in Biomedicine, ch. 8, pp. 211–235. Springer, New York (2005)
3. Boninski, P., Przelaskowski, A.: Content-based indexing of medical images for digital radiology applications. In: Pietka, E., Kawa, J. (eds.) Information Technologies in Medicine, pp. 113–120. Springer, Berlin (2008)
4. Cios, K.J. (ed.): Medical Data Mining and Knowledge Discovery. Physica-Verlag, Heidelberg (2001)
5. Corcho, O., Gomez-Perez, A.: Evaluating knowledge representation and reasoning capabilities of ontology specification languages. In: Proc. of the ECAI 2000 Workshop on Application of Ontologies and Problem Solving Methods (2000)
6. Fernandez-Lopez, M., Gomez-Perez, A.: Overview and analysis of methodologies for building ontologies. The Knowledge Engineering Rev. 17(2), 129–156 (2002)
7. Hawkins, J., Blakeslee, S.: On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machine (2004) (Polish translation: Helion, 2006)
8. Koczkodaj, W.W., Przelaskowski, A., Szopinski, K.T.: Medical knowledge mining from image data - synthesis of medical image assessment for early stroke detection. Machine Graphics and Vision 19(1) (2010) (in press)
9. Kulikowski, J.L.: Relational approach to structural analysis of images. Machine Graphics and Vision 1(1/2), 299–309 (1992)

10. Kulikowski, J.L.: From pattern recognition to image interpretation. Biocybernetics and Biomedical Engineering 22(3/4), 177–197 (2002)
11. Kulikowski, J.L.: Description of irregular composite objects by hyper-relations. In: Wojciechowski, K., et al. (eds.) Computer Vision and Graphics. Int. Conference ICCVG 2004, Proceedings, pp. 141–146. Springer, The Netherlands (2004)
12. Kulikowski, J.L.: The role of ontological models in pattern recognition. In: Kurzynski, M., et al. (eds.) Proc. of the 4th International Conference on Computer Recognition Systems CORES? 2005, pp. 43–52. Springer, Berlin (2005)
13. Kulikowski, J.L.: Interpretation of medical images based on ontological models. In: Rutkowski, L., et al. (eds.) ICAISC 2006. LNCS (LNAI), vol. 4029, pp. 919–924. Springer, Heidelberg (2006)
14. Kulikowski, J.L.: Structural image analysis based on ontological models. In: Kurzynski, M., et al. (eds.) Computer Recognition Systems, vol. 2, pp. 43–52. Springer, Berlin (2007)
15. Kulikowski, J.L.: Pattern Recognition driven by domain ontologies. In: Kurzynski, M., et al. (eds.) Computer Recognition Systems, vol. 3, pp. 43–52. Springer, Berlin (2009)
16. Kulikowski, J.L.: Logical inference based on incomplete and/or fuzzy ontologies. In: Krol, D., Nguyen, N.T. (eds.) Intelligence Integration in Distributed Knowledge Management. Information Science Reference, Hershey, ch.1, pp. 1–16 (2009)
17. Ogiela, M.R., Tadeusiewicz, R.: Picture languages in machine understanding of medical visualization. Machine Graphics and Vision 14(1), 71–82 (2005)
18. Penrose, R.: Shadows of the Mind. A Search for the Missing Science of Consciousness. Oxford University Press, Oxford (1994) (Polish translation: Zysk i S-ka Wyd. Poznan, 2000)
19. Tadeusiewicz, R., Ogiela, M.R.: Medical image understanding technology. Springer, Berlin (2004)

Unsupervised, Fast and Precise Recognition of Digital Arcs in Noisy Images*

Thanh Phuong Nguyen[1], Bertrand Kerautret[1],
Isabelle Debled-Rennesson[1], and Jacques-Olivier Lachaud[2]

[1] ADAGIo team, LORIA, France
{nguyentp,kerautret,debled}@loria.fr
[2] LAMA, University of Savoie, France
jacques-olivier.lachaud@univ-savoie.fr

Abstract. In image processing and pattern recognition, the accuracy of most algorithms is dependent on a good parameterization, generally a computation scale or an estimation of the amount of noise, which may be global or variable within the input image. Recently, a simple and linear time algorithm for arc detection in images was proposed [1]. Its accuracy is dependent on the correct evaluation of the amount of noise, which was set by the user in this former version. In the present work we integrate a promising unsupervised noise detection method [2] in this arc recognition method, in order to process images with or without noise, uniformly distributed or variable within the picture. We evaluate the performance of this algorithm and we compare it with standard arc and circle detection methods based on extensions of the Hough transform.

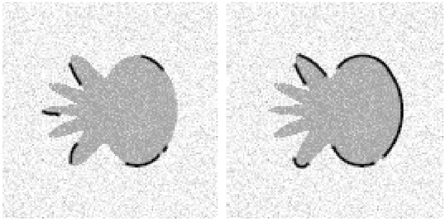
1 Introduction

Since the last three decades, the problem of arc and circle detection is intensively studied in the computer vision and pattern recognition community. Even if important progresses have been made from the classical generalized Hough Transform, current methods are always dependent on a good choice of parameters. As shown in Fig. 1, a correct choice of parameter is necessary to get good quality in the arc detection (method of [1]). Furthermore, the complexity of several detection methods is very dependent on their parameterization, which may forbid their use in real time applications (see example (f) of

The arc and circle detection gives important information for detecting specific objects in digital images or for shape recognition. We can classify the existing methods into two main groups: image and contour-based approaches. The first group contains the methods that detect directly arcs and circles through some well known techniques of transformation in images (Hough or Radon transform). The generalized Hough transform is widely used as a basic tool for solving this problem[3–5]. These methods try to transform the input image into the parameter space and the circle candidates are detected as peaks in this space. However,

* This work is supported by ANR in the framework of the GeoDIB project (ANR-06-BLAN-0225).

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 59–68, 2010.



a) b)

Fig. 1. Detected arcs at different widths 2 (a) and 3 (b)

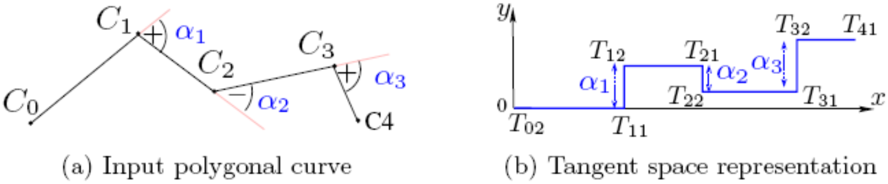
these methods require a lot of computation and memory space, which hinders their efficiency when detecting large arcs or circles. In addition, they depend on a threshold for both accepting short arcs and eliminating noises. The second group of methods contains the ones that recognize arcs and circles on extracted contours [1, 6]. Although some of them can process damaged contours [1], all require user parameterization indicating the expected amount of noise. The approach of Roussillon *et al.*allows to perform arc recognition in linear time however it requires one of three constraints (see [7] for more detail).

The main contribution of this work is to propose a simple, efficient and unsupervised contour based algorithm for arc and circle detection. The method is based on the linear time supervised approach [1] and exploits an original and simple automatic noise detector based on discrete geometry [2]. It can process indifferently perfect images, or damaged by a uniformly distributed or variable noise. In addition, it is more efficient than traditional and improved methods based on Hough transform. The paper is organized as follows. Section 2 introduces a linear method for arc recognition. The unsupervised noise detection is presented in section 3. In section 4, we present a framework for arc segmentation along noisy curves. Several experiments and comparisons are given in section 5.

2 Linear Method for Arc Recognition and Segmentation

The arc recognition algorithm proposed in [1] is based on some properties of arcs in the tangent space representation introduced by Latecki *et al.* [8]. The discrete circles are seen through the arithmetic definition of Andres [9] that defines a digital circle as a sequence of points included in a ring.

Tangent space representation for arc recognition. The main idea of this representation is to transform the polygon $C = \{C_i\}_{i=0}^{n}$ of $\mathbb{Z}^2$ by using lengths and angles between the successive segments of this polygon (see Fig. 2). We denote by α_i the angle $\angle(\overrightarrow{C_{i-1}C_i}, \overrightarrow{C_iC_{i+1}})$. If C_{i+1} is on the right hand side of $\overrightarrow{C_{i-1}C_i}$ then $\alpha_i > 0$, otherwise $\alpha_i < 0$. Some properties of a set of sequential

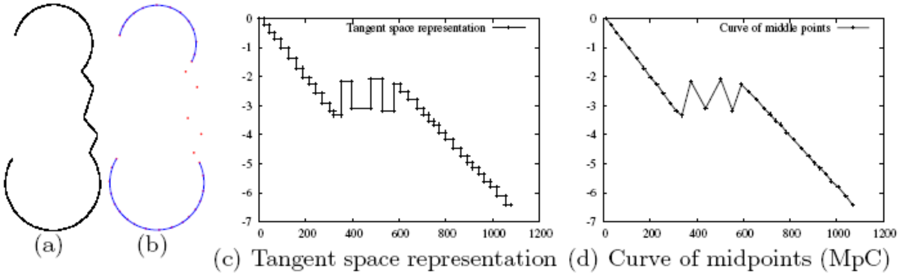


(a) Input polygonal curve (b) Tangent space representation

Fig. 2. Tangent space representation

chords of a circle in the tangent space have been established in [1]. They are summarized in the following property 1.

Property 1. Let $C = \{C_i\}_{i=0}^{n}$ be a polygon, $\alpha_i = \angle(\overrightarrow{C_{i-1}C_i}, \overrightarrow{C_iC_{i+1}})$ such that $\sin\alpha_i \simeq \alpha_i$, $0 < i < n$. The length of C_iC_{i+1} is l_i, for $i \in \{1, \ldots, n\}$. We consider the polygon $T(C)$, that corresponds to its representation in the modified tangent space, constituted by the segments $T_{i2}T_{(i+1)1}$, $T_{(i+1)1}T_{(i+1)2}$ for i from 0 to $n-1$. $MpC = \{M_i\}_{i=0}^{n-1}$ is the midpoint set of $\{T_{i2}T_{(i+1)1}\}_{i=0}^{n-1}$. So, C is a polygon that approximates a circle or an arc of circle if and only if $MpC = \{M_i\}_{i=0}^{n-1}$ is a set of collinear points.



(a) (b) (c) Tangent space representation (d) Curve of midpoints (MpC)

Fig. 3. Arc segmentation results on a digital curve: (a) input curve, (b) vertex of the approximated polygon and result of arc segmentation

Blurred Segment of width ν [10]. It is a set of points that is included in a band of width ν. It can thus represent a discrete straight segment while taking into account the amount of noise contained in the initial data (according to ν). This notion is used for both the polygonalisation step (ν_1) and for the collinearity evaluation (ν_2).

The main steps of the algorithm are illustrated on Fig. 3: (a,b) polygonalization step, (d) associated tangent space representation (d). The midPoints curve (e) highlights arc circles in its straight parts (c). The algorithm is detailed in [1].

3 Unsupervised Noise Detection

In the work presented in [2], the authors propose a simple and efficient approach for detecting noise along discrete contours. We present here briefly the main ideas and refer the reader to the preceding reference for more details.

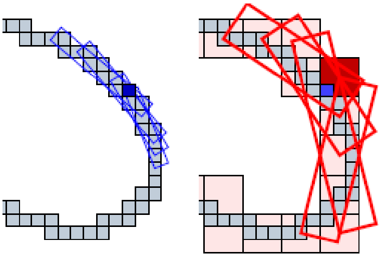
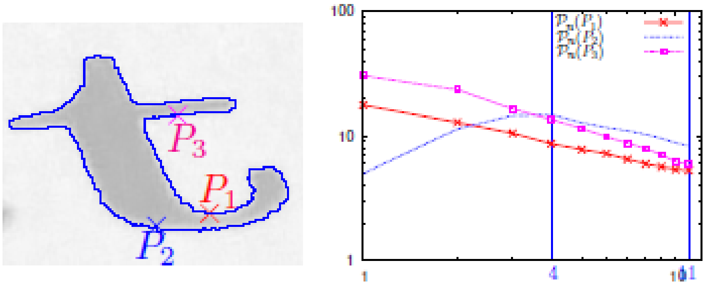


Fig. 4. Illustration of maximal segments on (a) an initial contour C, and (b) on its subsampled contour $\phi_3(C)$

The noise estimation is linked to the asymptotic properties of the *maximal segment* primitive. We recall first that a standard digital straight line (DSL) is a 4-connected digital set $\{(x, y) \in \mathbb{Z}^2, \mu \leq ax - by < \mu + |a| + |b|\}$, all parameters being integers. A *digital straight segment (DSS)* is a 4-connected piece of DSL. If we consider a 4-connected path C, a *maximal segment* M is a subset of C that is a DSS and which is no more a DSS when adding any other point of $C \backslash M$.



a) b)

Asymptotic properties of maximal segments. The presence of noise is related to the discrete lengths of maximal segments. Let us denote by $\text{Dig}_h(X)$ the Gauss digitization with step h of a shape X. Let p be some point on the boundary of X. We denote by (L_j^h) the discrete lengths of the maximal segments, defined around $\text{Dig}_h(X)$, and which cover p. It can be shown that if X is a simply connected shape in $\mathbb{R}^2$ with a piecewise C^3 boundary ∂X and U an open connected neighborhood of p, then we have:

$$\text{if } U \text{ is strictly convex or concave, then } \Omega(1/h^{1/3}) \leq L_j^h \leq O(1/h^{1/2}) \quad (1)$$

$$\text{if } U \text{ has null curvature everywhere, then } \Omega(1/h) \leq L_j^h \leq O(1/h) \quad (2)$$

Since in practice, it is not possible to obtain the asymptotic digitizations of the initial shape O with finer and finer grid steps h, a solution is to consider the subsampling $\phi_i(O)$ with increasing covering pixel sizes $i \times i$ for $i = 2, .., n$. Then, we can consider the discrete lengths $(L_j^{h_i})$ of the maximal segments on the subsampled shapes with the increasing sequence of digitization grid steps $h_i = ih$ (see

For a given subsampling size i, the average discrete length of all the maximal segments containing the subsampled pixel is denoted as $\overline{L}^{h_i}$.

Multiscale profile and noise detection. The multiscale profile $\mathcal{P}_n(P)$ of a point P is defined as the sequence of samples $(X_i, Y_i) = (log(i), log(\overline{L}^{h_i}))_{i..n}$. Some multiscale profiles are illustrated on the real contour of the following floating figure. According to inequality (1) (resp. (2)), if P is located on curved (resp. flat) part, the slope of an affine approximation of the multiscale profile should be in $[-\frac{1}{2}, -\frac{1}{3}]$ (resp. $[-1, -\frac{1}{2}]$). Since for noisy contour parts the preceding properties are not valid, it is used to detect such a part. A threshold t_m is given to determine the *meaningful scale* defined as a pair (i_1, i_2), $1 \leq i_1 < i_2 \leq n$, such that for all i, $i_1 \leq i < i_2$, $\frac{Y_{i+1}-Y_i}{X_{i+1}-X_i} \leq t_m$. For the example of Fig. 5, the meaningful scales of the points P_1, P_2 and P_3 are respectively equal to $(1, 11)$, $(4, 11)$ and $(1, 11)$.

From the previous meaningful scale (i_1, i_2) , $i_1 - 1$ defines the *noise level* of a point P (denoted as $\eta(P)$). Experimentally the threshold value $t_m = 0$ gives best results both on curved or flat noisy parts. Fig. 5 shows some results obtained on various shapes (a,b). The noise level is well detected on both synthetic and real images. The curved/flat part areas of a shape can also be detected by thresholding the slope of the multiscale profile on the meaningful scale (see the multiscale of the point P_3 of the previous floating figure). Fig. 5 (c) shows such a detection (curved area represented by a thick contour).

4 A Framework for Arc Recognition along Noisy Curves

As mentioned in Section 2, the arc detection algorithm uses a parameter ν_1 to take into account the amount of noise in the polygonalization step. This parameter is adjusted manually to obtain a representative polygon that allows a good arc recognition. As mentioned in [1], an interesting question is raised: *"For each noisy curve, how can we choose the value of ν_1 to obtain the best result?"*. This section proposes a framework to obtain a parameter-free method.

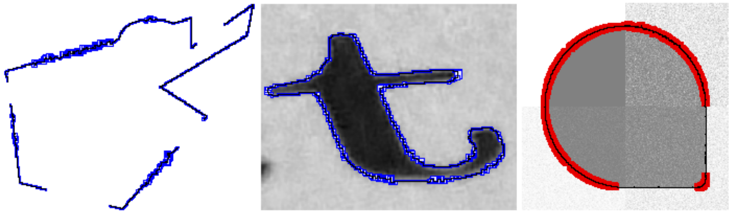


Fig. 5. Noise level detection (a,b). For each point P the noise level is represented by a centered box of size $\eta(P) + 1$. (c) Curved/flat areas detection from meaningful scales (curved areas represented by a thick contour).

4.1 Polygonalization Adapted to Noisy Curves

We propose here two solutions for the arc recognition on noisy curves. Both solutions construct an approximating polygon by taking into account the amount of noise of the discrete contour. We have considered different hypotheses on the noise distribution. The first one considers the hypothesis of uniform distribution while the second one is an adaptive solution which take into accounts locally the amount of noise of the discrete contour.

Algorithm 1. Polygonalization based on unsupervised noise detection

```
Data: $C = \{C_i\}_{i=0}^{n}$ digital curve, $\eta = \{\eta_i\}_{i=0}^{n}$ noise information,
      $uniformNoise$- $true$ if uniform noise distribution, $false$ otherwise
Result: P-approximated polygon
begin
    b ← 0; Add $C_b$ to $P$;
    if $!uniformNoise$ then
        while $b < n$ do
            Use [10] to recognize $\{C_b, \ldots, C_e\}$ as blurred segment of width $\eta_b$;
            b ← e ; Add $C_b$ to $P$;
    else
        $\bar{\eta}$ ← mean value of $\eta = \{\eta_i\}_{i=0}^{n}$;
        while $b < n$ do
            Use [10] to recognize $\{C_b, \ldots, C_e\}$ as blurred segment of width $\bar{\eta}$;
            b ← e ; Add $C_b$ to $P$;
end
```

These solutions are based on the unsupervised noise detector presented in Section 3. This method gives noise information at each point of the noisy input curve. The local noise estimation is used to construct a polygonal approximation of the curve, adapted to the local amount of perturbation. These two processes of polygonalization are described by Algorithm 1. It considers two possible options for the noise distribution:

Uniform noise distribution. In this case the approximated polygon is constructed by using a constant width value in the blurred segment based polygonalization. This width value is defined as the mean noise level $\bar{\eta}$ obtained on the whole input curve.

Non-uniform noise distribution. Here the construction of the approximated polygon is obtained by modifying the width of the blurred segments from each new segment recognition. The width is chosen according to the noise level η of the first point added to the blurred segment.

Independence with respect to starting point. To do so, the information related to flat/curved areas determines the better way to start the polygonalization. Indeed we compute the largest interval (f_b, f_e) $b, e < n$ of the contour that is a flat area. Then, if the contour is closed, we apply Algorithm 1 with a circular permutation of size e of the initial contour. Otherwise Algorithm 1 is called twice with the partial contours $C'_{e,n}$ and $C'_{0,b}$.

4.2 Arc Recognition along Noisy Curves

Algorithm 2 presents a parameter-free approach for arc segmentation along noisy curves. As it is shown in [1], the error bound of arc recognition depends on the width ν_2 of the blurred segment recognition and on the maximal admissible variation angle α_{max} in the tangent space. To guarantee the quality of segmentation, we fix these parameters to $\nu_2 = 0.25$ and $\alpha_{max} = \frac{\pi}{4}$ in Algorithm 2 (see also section 7.4 of [1] for more details and for the α_{max} setting). Therefore, the proposed method is in practice parameter-free when we work with noisy curves,

Algorithm 2. Arc segmentation along a noisy digital curve

```
Data: $C = \{C_i\}_{i=0}^{n}$ noisy digital curve
Result: ARC- sequence of extracted arcs
begin
    $N \leftarrow \{N_i\}_{i=0}^{n}$ noise information determined by [2] (see Section 3);
    $ARC \leftarrow \emptyset$; Use Algorithm 1 to polygonalize $C$ in $P = \{P\}_{i=0}^{m}$;
    Represent $P$ in the tangent space by $T(P)$ (see Section 2);
    Determine the midpoint set $MpC = \{M_i\}_{i=1}^{n}$ (see Section 2);
    Use [10] to polygonalize $MpC$ into a sequence $S = \{S\}_{i=0}^{k}$ of blurred
    segments of width 0.25;
    for $i$ from 0 to $k-1$ do
        $\{M_j\}_{j=b}^{e}$: sequence of points of $MpC$ that corresponds to $S_iS_{i+1}$;
        $C'$: part of $C$ that corresponds to $S_iS_{i+1}$;
        $isArc \leftarrow true$;
        for $i$ from $b$ to $e-1$ do
            if $M_{i+1}.y - M_i.y > \frac{\pi}{4}$ then $isArc \leftarrow false$
        if $isArc$ then Add $C'$ to $ARC$
end
```

thanks to the unsupervised noise detection. It uses no thresholding and automatically provides detected arcs in noisy curves.

5 Experimental Evaluation

The first experiments were performed on the noisy synthetic shapes of Fig. 6. Gaussian noise was added globally in images (a,b) and specifically to each quadrant in images (c,d). The user has only to choose between a uniform or a variable noise distribution to obtain a more precise arc detection (boolean parameter $uniformNoise$). Another experiment on a real image from a 10 million pixels digital camera is presented on Fig. 7. All the experiments presented here were obtained on a *Mac OS X*, $2.8GHz$, $4GB$ $DDR3$ *Intel Core Duo*. The contours were extracted from a simple threshold set to 128. Fig. 8 presents some others experiments on real pictures, with also a comparison with three methods based on the Hough transform: CHT [3], MHT [4] and FHT [5]. For all of these experiments the source image was thresholded by a value defined in the HSB color space. The numerous parameters μ_C, μ_M, and μ_F were respectively associated to $(radiusMin, radiusStep, nbCircles)$, $(minRadius, maxRadius)$, $(gradTreshold, radius, radiusVariation)$. Note that other parameters were not mentioned here, like for FHT and can have some influence in the result (α and β see [5]).

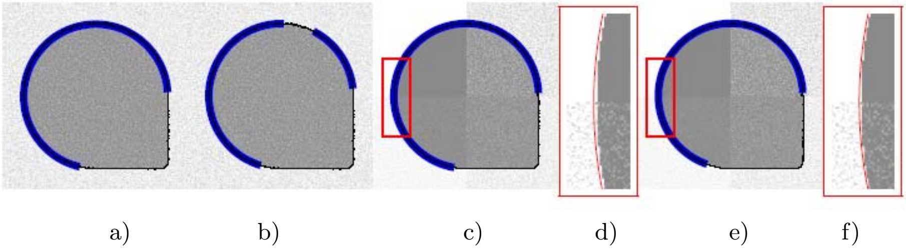


Fig. 6. a) global 1.07s 2228 points; b) adaptive 0.98s 2228 points; c) global 0.96s 2228 points; e) adaptive 0.88s 1926 points.

Arcs detection from the global noise based approach (a,d) and the adaptive approach (b,e) (image size 512x512). (d,f) close-up view of (c,e).

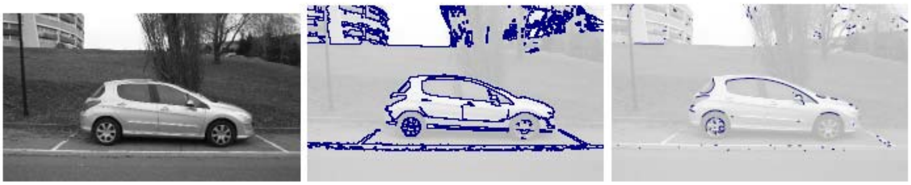
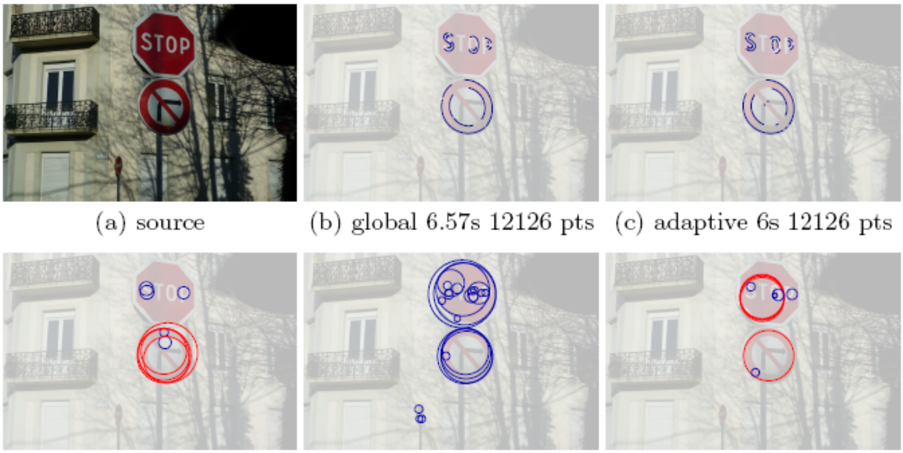


Fig. 7. Arc detection with our method on an image of a car (size 4000x2672 pixels)



(a) source (b) global 6.57s 12126 pts (c) adaptive 6s 12126 pts

(d) CHT [3](1024x684)
$\mu_C = (70, 2, 25)$ 1m19s
$\mu_C = (5, 1, 20)$ 1m0s

(e) MHT [5](1024x684)
$\mu_M = (10, 190)$ 2.0s

(f) FHT [4] (4000x2672)
$\mu_F = (200, 330, 100)$1m27s
$\mu_F = (170, 50, 100)$ 4m26s

Fig. 8. Application of our method on a real picture (size 4000x2672 pixels) with the possible values for *uniformNoise* (a-c), and comparison with three methods based on Hough transform (d-f)

6 Conclusion and Perspectives

We have presented a new approach for arc segmentation of digital curves in noisy images by combining it with an unsupervised noise detector. The proposed methods lead to an efficient arc detector in images, especially in comparison with methods based on the Hough transform, which require both large memory and execution time. Moreover the proposed method is not dependent to the need to set a specific parameter. In the future, we plan to integrate the detection of curved zone in noisy curves of [2] (see also figure 5(c)) as a preprocessing step to enhance the robustness of the arc detector.

References

1. Nguyen, T.P., Debled-Rennesson, I.: A linear method for segmentation of digital arcs. Technical report, LORIA (2010), http://www.loria.fr/~nguyentp/pubs/techreport_arcsegmentation.pdf
2. Kerautret, B., Lachaud, J.O.: Multi-scale analysis of discrete contours for unsupervised noise detection. In: Wiederhold, P., Barneva, R.P. (eds.) IWCIA 2009. LNCS, vol. 5852, pp. 187–200. Springer, Heidelberg (2009)
3. Kimme, C., Ballard, D., Sklansky, J.: Finding circles by an array of accumulators. Short Communications Graphics and Image Processing 18, 120–122 (1975)
4. Davies, E.: A modified hough scheme for general circle location. Pattern Recognition Letters 7, 37–43 (1984)

5. Rad, A.A., Faez, K., Qaragozlou, N.: Fast circle detection using gradient pair vectors. In: Digital Image Comp.: Techniques and Applications, pp. 879–887 (2003)
6. Coeurjolly, D., Gérard, Y., Reveillès, J.P., Tougne, L.: An elementary algorithm for digital arc segmentation. Discrete Applied Mathematics 139, 31–50 (2004)
7. Roussillon, T., Tougne, L., Sivignon, I.: On three constrained versions of the digital circular arc recognition problem. In: Brlek, S., Reutenauer, C., Provençal, X. (eds.) DGCI 2009. LNCS, vol. 5810, pp. 34–45. Springer, Heidelberg (2009)
8. Latecki, L., Lakamper, R.: Shape similarity measure based on correspondence of visual parts. IEEE Transactions on PAMI 22, 1185–1190 (2000)
9. Andres, E.: Discrete circles, rings and spheres. Comp.&Graphics 18, 695–706 (1994)
10. Debled-Rennesson, I., Feschet, F., Rouyer-Degli, J.: Optimal blurred segments decomposition of noisy shapes in linear time. Computers & Graphics 30 (2006)

The Role of Sparse Data Representation in Semantic Image Understanding

Artur Przelaskowski

Institute of Radioelectronics, Warsaw University of Technology
Nowowiejska 15/19, 00-665 Warsaw, Poland
arturp@ire.pw.edu.pl

Abstract. This paper discusses a concept of computational understanding of medical images in a context of computer-aided diagnosis. Fundamental research purpose was improved diagnosis of the cases, formulated by human experts. Designed methods of soft computing with extremely important role of: a) semantically sparse data representation, b) determined specific information, formally and experimentally, and c) computational intelligence approach were adjusted to the challenges of image-based diagnosis. Formalized description of image representation procedures was completed with exemplary results of chosen applications, used to explain formulated concepts, to make them more pragmatic and assure diagnostic usefulness. Target pathology was ontologically described, characterized by as stable as possible patterns, numerically described using semantic descriptors in sparse representation. Adjusting of possible source pathology to computational map of target pathology was fundamental issue of considered procedures. Computational understanding means: a) putting together extracted and numerically described content, b) recognition of diagnostic meaning of content objects and their common significance, and c) verification by comparative analysis with all accessible information and knowledge sources (patient record, medical lexicons, the newest communications, reference databases, etc.).

Keywords: Computational intelligence, image understanding, sparse image representation, nonlinear approximation, semantic information theory.

1 Introduction

A process of radiological interpretation generally includes the understanding of medical image content resulting in recognition of possible pathology symptoms, most often called detection, and assessment of comprehensive image information in a context of current clinical case-knowledge. It involves image-based detection of disease, defining disease extent, determining etiology of the disease process, assisting in designing of the clinical management plans for the patient, based on imaging findings, and following response to the therapy. This exclusive work of humans needs computer-based support mostly because of limited accuracy of diagnosis.

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 69–80, 2010.

1.1 Error in Radiology

Errors occur frequently in the practice of radiology. Many years ago L.H. Garland noticed that a substantial degree of observer error was prevalent in radiologic interpretation – *not only should clinicians recognize their own errors; they should admit them* [1]. Approximately 4% of radiological interpretations rendered by radiologists in their daily practice contain errors [2]. The rate of false negatives approximates 30% [3]. Moreover, such high rate of observer errors is stable across tens of years [3, 4].

Generally two kinds of errors are distinguished: perceptual errors when recorded image features were not appreciated and cognitive (reasoning or interpretation) errors whereas image features, though appreciated, led to wrong conclusions. Perceptual error occurs four times more frequently than cognitive error. Underreading error occurs five times more frequently than overreading error [4].

Supposed set of error causes includes:

- diagnostic knowledge shortage, i.e. a shortage of reliable gold standards (procedures, objective and unique criteria, pathology and normality patterns), shortage of formal knowledge (evidences, models, lexicons, ontologies), shortage of specific knowledge or experience (reference datasets, patterns – any lesion manifestation is significantly diversified across scale, form, number, size, shape and localization), lack of consultation conditioning;
- recognition difficulty because of limitation of imaging technology, interpretation conditioning, diverse case-specificity, uncertainty in decision making (incomplete information available, ambiguous rationales);
- general human factor problem, i.e. human being perspective (fatigue, subjectivity, observer variations), imperfect human perception (blind spot, pseudovision, attention, first sight, image scanning order, intuition, happy eye syndrome), unexplained human equation in interpretive process (partly visual and partly psychological, with perceptual and intellectual aspects);
- environmental problems, i.e. limited technical conditioning of imaging (inadequate radiological examination, the abnormality is outside the area of primary examination), poor communication, poor work organization (ergonomics, equipment quality, clinical history access, prior examinations access), misdirected search.

1.2 Computer-Based Improvements of Diagnosis

Among methods of reducing radiological errors, the following procedures were suggested: supporting better quality of images, i.e. more sensitive and straightforward imaging technologies, suitable time spent interpreting studies or optimal volume of cases interpreted per unit of time, comparison with previous radiological studies, reasonable use of clinical history (before or after initial interpretation?), double interpretation or consultations, care, training and experience, education, teaching of normal anatomic and physiologic variants, elucidation and correction of the factors involved (retrospective evaluation of missed cases, group discussion) [5–8], etc.

Extensions of these postulates require computer technologies and information systems. According to many experiences, digital technology can potentially improve screening and even diagnostic procedures by providing compatibility for computer-assisted detection of lesions and by permitting flexible manipulation of gray-scale and edge enhancement of the images presented to the radiologist [9]. Real challenges formulated for compute assistance of image diagnosis include the following effects, results and general inputs:

- effects of perception enhancement of image content (semantically-driven image processing), extraction of subtle or visually hidden information (semantically-driven image analysis), automatic indications of abnormalities – diagnostic content understanding and diagnostic interpretation of the objects;
- results for
  - perception – increased sensitivity of scanning procedures,
  - cognition – consultant opinion of computational intelligence,
  - decision – assistance in careful synthesis of evaluation and conclusion;
- general inputs as follows:
  - formalized, structured hierarchy of knowledge domain – ontology for standardized rating criteria and semantic-driven models of computational intelligence,
  - objectivization and systemization of interpretation process (i.e. criteria, comparisons, rules, reports, evaluation factors),
  - straightforward, consensus definition of diagnostic decision rules (scales, recommendations etc.),
  - reference case study through content-based image retrieval.

Computer-based aiding tools are able to storage, index and retrieve, communicate, select, emphasize, analyze, recognize, understand and visualize image-based diagnostic content. Because the initial and key constituents of diagnosis are the true detection and defining of the disease, full understanding and accurate assessment of image content including semantics of recognized objects, mutual relations and accessible information complements is a key issue of successful exploitation of imaging capabilities in diagnosis.

2 Concepts of Image Understanding

Generally, image understanding is originated in semantic information theory[1] and means recognized essence of diagnostic image information. It means additional perspectives on the same image, taking into account the meaning and significance of its content.

According to research of R. Tadusiewicz and M. Ogiela [10–12], the human understanding process is always based on the physician knowledge opposite to typical image processing and analysis methodology, which is ever data-driven procedure. The task performed by the doctors mind is all the time knowledge-based reasoning. Thus, automatic image understanding investigates the meaning

[1] http://plato.stanford.edu/entries/information-semantic/

of the fact that visible objects have certain features and reasons consequences arising from the fact that objects may be assigned to certain selected classes. The proposed solutions imitate a natural way of thinking of a human expert by linguistic description of the meritorious content of the image allowing its semantic representation without an aprioristic definition of a number of the recognized classes. The created linguistic description is of a structural character and allows analyzing the meaning as detailed as needed for classification or indexation purposes. Therefore, during the linguistic analysis of medical patterns the problem of generalization of features of a selected image to obtain semantic content description is solved [13].

Semantic content descriptors, extractors and enhancements are fundamental for providing automatic interpretation of diagnostic image information. High-level analysis and recognition include the use of computationally intelligent techniques based on effective data representation and models adjusted with help of functional analysis tools, atoms of harmonic analysis, approximation theory methods and human visual models for image perception and interpretation. The relationship between image components, objects and patterns taking into account occurrence context plays fundamental role for diagnosis. It is modeled, adjusted and classified according to representative diagnostic knowledge gained from reliable formal and informal sources. Computational image interpretation are expected to be additional, semantic eyes of diagnosis.

More generally, in a heterogeneous world of ubiquitous information flow semantic technologies allow a flexible and seamless integration of applications and data sources. They provide an intelligent access with understanding of context-based content, give answers and generate knowledge including as objective as possible object description. Semantic annotation of images is a key concern for the newly emerged applications of networked multimedia. Automatic understanding to improve machine interpretation of the images is actually demanding challenge. Quality image estimates, numerical accuracy estimates and semantic information metrics reflect automatic understanding of the image content. Numerical descriptors of image content resonant to semantic image extent are designed according to structured knowledge and following expert requirements. The expected results are computational models of image semantics.

2.1 Conditioning of Cognitive Resonance

Key problem of a semantic gap between the low level numerical descriptors and the high level image interpretation in diagnostic categories is still challenging problem of computer-aided diagnosis. Because low level descriptors cannot be uniquely associated with any meaningful label unless explicitly declared or derived as the outcome of a classification procedure, retrieval or automatic recognition and evaluation based on knowledge level constructions is a non-trivial task to achieve in general.

Suggested solution is cognitive resonance of selective sparse descriptors with structured or empirically established knowledge platform. Considered methodology includes estimation of semantic sparse representation of an image that is optimized with medical knowledge platform. Thanks that following classification of extracted semantic components is simplified and designed according to specified diagnostic categories.

Principal issue of image understanding is image content modeling based on semantic sparse representation designed with the following conditions:

- initial approximation of diagnostic content with the methods of nonlinear target estimation, nontrivial segmentation, specific denoising, etc.;
- as sparse as possible image data representation arranged locally or globally according to diagnostic reasons;
- components of the model defined according to imaging reality and interpretation task;
- resulting representation susceptible to flexible, reliable and specific numerical description of diagnostic image content, i.e. meaning of local regions, detailed structures, textures and global image features;
- semantic adjusting of the representation reasoned in medical knowledge platform by:
  - characteristics of observer possibilities and limitations due to image perception and understanding,
  - reliable observer performance characteristics – a methodology of interpretation: rules and protocols of effective diagnosis,
  - formalized, hierarchical medical knowledge platform: objective taxonomies and complete ontologies,
  - informal knowledge extent basing on empirical heuristic procedures of reference pattern analysis;
- sparse data classification in diagnostic categories.

3 Representation Based Approach as a Method

The problem of data representation is one of the most critical issues concerning the realization of intelligence machines which are able to solve real life problems [14]. Adjustment of signal or source data representation to application requirements is a clue problem of many theories and algorithmic solutions. General purpose is successful separation of a signal content to effectively manipulate it.

Adjustment of image representation to automatic diagnosis requirements was optimized with criteria of sparsity of semantically modeled content. Proposed outline of medical image understanding for computer-aided diagnosis purposes uses medical knowledge platform and consists of:

- estimation of target signal with nonlinear approximation procedure based on approximants adjusted to whole image diagnostic model (including morphology, important structures, area of interests, noise and artifacts characteristics, etc.);
- sparse decomposition of local texture characteristics based on
  - fixed patch cover according to diagnostic knowledge guidelines,
  - multi-component analysis with redundant dictionaries,

  - multiscale, local basis of dictionary adjusted to local diagnostic model for regions of interests;
- simplified supervised classification of local textures based on
  - diagnostic categories of specific application,
  - features extracted from semantic sparse representation of the image,
  - training sets of representative pathology patterns and control groups.

Formal description of selected items possible realization was shortly outlined.

3.1 Formal and Informal Medical Knowledge Platform as a Method

More formal medical knowledge extraction like ontologies should be completed with experimentally selected pathology patterns verified according to reference database paradigm. Primary goal of ontology is to represent effectively a domain knowledge, adequately and exhaustively define relevant concepts, object characteristics and relationships between them, to provide a common, standardized vocabulary comprehensible by humans and machines by which users and computer systems can communicate. Thus, ontology means systemization, objectivization and verification of knowledge base to formulate the model populated with concept's instances. Such structured model constitutes standard diagnostic knowledge database. Ontologies are the foundation of the Semantic Web, where integration and interoperability of heterogeneous sources of information is needed. Ontologies also form the basis foundation of evidence-based-medicine and standardization efforts.

An implementation of mammographic onthology [15] is exemplary realization of formalized medical knowledge platform, useful for automatic understanding of mammograms according to proposed image understanding methodology.

3.2 Sparse Data Representation

Sparse data representation is considered in a context of analyzing a given signal f using the local dictionary $\mathcal{D}$ of basis (atoms) well approximating investigated features of f. Sparse signal expansion in $\mathcal{D}$ means that only a few atoms of $\mathcal{D}$ are active to describe f.

Sparsity leads to efficient data modeling and separation, estimations, coding, dimensionality reduction, feature extraction and pattern recognition. For instance, respective data representation influences the generalization error of kernel based learning machines like Support Vector Machines (SVM) for classification [16]. In case of sparse data representation, the generalization capacity of an SVM trained by using polynomial or Gaussian kernel functions is equal to the one of a linear SVM. It means that sparse data representations reduce the generalization error as long as the representation is not too sparse, as in the case of very large dictionaries.

Sparse texture analysis is used for semantic component selection. Identification of dominant morphological ingredients is the most optimistic step for image analysis. For the purpose of analysis, extracted texture characteristics may be useful for specific object recognition. Defining object class of interests is derived from semantic problem description. Numerical data analysis is used for sensing compliment of dominant, recognized and identified basic content. But computer aided diagnosis (CAD) is directed mostly to subtle signatures of pathology in a context of general content characteristics. Thus formalized medical knowledge is used for dominant component recognition but empirical knowledge from reference database is used for subtle but key compliment of diagnostic image understanding for accurate interpretation. In other words, probable anomalies (ROI) detection is based on multi-component procedure with local but exhausting analysis based on sparse patch representation. Detected subtle lesions are verified and interpreted with help of differential texture analysis completed by supervised empirical procedure.

3.3 Nonlinear Approximation

The fundamental problem of approximation theory is to resolve a possibly complicated target function by simpler, easier to compute basis functions called the approximants. Formally, an approximation process can be simply defined in a Hilbert space with inner product $< .,. >$ and norm $\left\|.\right\|_H$. Let $\{\varphi_i\}_{i=1,2,...}$ be an orthonormal basis for complete $\mathbf{H}$. Each signal $f \in \mathbf{H}$ can be decomposed in this basis $f = \sum_{i=1}^{+\infty} a_i \varphi_i$ with the coefficients of orthogonal expansion $a_i =< f, \varphi_i >$.

In case of nonlinear approximation we use the nonlinear space $\mathbf{A}_N$ for expression of $\overset{\approx}{f} \in \mathbf{H}$, where $\mathbf{A}_N$ replaces linear space of dimension N, typically used for discrete signal representation. Signal approximation $\overset{\approx}{f}$ is defined as $\overset{\approx}{f}_{(N)} = \sum_{i \in \Lambda} a_i \varphi_i$, where $\Lambda \subset \mathbf{N}$ is a finite set of indexes with the cardinality $\#\Lambda = M \ll N$ for sparse representation. M terms are chosen depending on the meaningful features of approximated f.

Nonlinear approximation error $\overset{\approx}{\varepsilon}{}_N^2(f) = \sum_{i \notin \Lambda} |a_i|^2$ is minimal and decays as M increases if Λ corresponds to the M vectors that best correlate to f, i.e. having the largest coefficients of the expansion $|a_i|$. For the set of indexes $\Lambda_r = \{i_k\}_{k=1,2,..,M}$ sorted according to decreasing order of the corresponding coefficients - $|a_{i_k} =< f, \phi_{i_k} > | \geq |a_{i_{k+1}}|$, $0 < k < M$ we have f approximation $\overset{\approx}{f}_{(M)} = \sum_{k=1}^{M} a_{i_k} \varphi_{i_k}$ with the error $\overset{\approx}{e}{}_N^2(f) = \sum_{k=M+1}^{+\infty} |a_{i_k}|^2$. The decay rate of $\overset{\approx}{\varepsilon}{}_N^2(f)$ as M increases is a measure of approximation efficiency. Consequently, the basis functions of approximation space should represent the most important, precisely characterized and distinguishable features of the target function, according to knowledge platform.

Among possible bases, wavelets are tailor-made for nonlinear approximation because of fast and simple computation, simplified characterization of approximation spaces based on almost unconditional function classes with controlled regularity and transparent strategies of basis selection and target function estimates. A nonlinear approximation in a wavelet orthonormal basis defines an adaptive grid that refines the approximation scale in the neighborhood of the signal singularities. If the target function is smooth on a region of interests a coarse scale of dyadic decomposition is used. For regions where the target function is less smooth, wavelet functions of higher resolution scales are used for the approximation. More accurate allocating terms in a nonlinear strategy depend on signal energy distribution across scales and subbands.

More efficient nonlinear image approximation may be constructed with scaled basis functions whose support shape can be flexibly adapted to the orientation and regularity of the object contours. It refers to non-separable wavelet kernels called 2D wavelets with anisotropic dilations, rotations and translations of mother function $\varphi_{m,n,\theta}(\cdot,\cdot)$ oriented by θ. The basic concepts of 2D wavelets use adaptive geometry-based approaches such as wedgelets (beamlets, platelets etc.), or directional frames such as ridgelets, curvelets, contourlets [17].

The best nonlinear approximation may be calculated by applying thresholding function with zeroing the expansion coefficients of magnitude less than the threshold value τ. For the source signal $s = f + \eta$ (with masking background η) we can estimate target function f by selection of the coefficients $a_i^{(s)}$ with thresholding function $d(\cdot)$ as follows: $\hat{f} = \sum_{i=1}^{N} d(a_i^{(s)})\varphi_i$.

The thresholds are matched adaptively considering coefficients distribution across scales, subbands and orientations keeping only wavelet domain transients coming from the target function. A wavelet thresholding is equivalent to estimating the signal by averaging it with a kernel that is locally adapted to the signal regularity.

3.4 Multi-component Data Representation

One of possible realization is morphological component analysis (MCA). It was designed to separate several components which have different morphologies through decomposition of the signal into building blocks [18]. MCA decomposition exploits morphological diversity of selected data features associated to respective atoms of used dictionary. Fundamental assumption is that for every signal target behavior to be separated, there exists in dictionary a basis enabling its sparse and as sparse as possible representation.

Let a given signal f be a sum of K components, i.e. $f = \sum_{k=1}^{K} f_k$, having different morphology μ_k and meaning Σ_k: $f_k(\mu_k, \Sigma_k)$ according to available *a priori* knowledge. We assume that a dictionary of basis $\mathcal{D} = \{\Phi_1, \ldots, \Phi_K\}$ exists such that for each k component f_k is satisfactorily sparse in respective Φ_k. It means that $\forall j \neq k, ||\Phi_k^T f_k||_0 < ||\Phi_k^T f_j||_0$ and $||\Phi_k^T f_k||_0 << ||f_k||_0$, where $||f||_0$ denotes the l_0 pseudo-norm of the vector (*de facto* the number of nonzero coefficients of f).

To make the problem solution more accurate and useful, two semantic components of lesion symptom f_l and normal background f_n were assumed. We have $f = f_l + f_n$ with heuristically and empirically determined bases Φ_l and Φ_n, respectively. The components f_l and f_n are estimated by solving the following constrained optimization problem:

$$\min_{f_l, f_n} \{||\Phi_l^T f_l||_1 + ||\Phi_n^T f_n||_1\} \quad \text{subject to} \quad ||\Sigma_f - \Sigma_{f_l} - \Sigma_{f_n}||_2 < \Sigma_\sigma \qquad (1)$$

where $\varSigma_\sigma$ is acceptable approximation of meaning $\varSigma_f$ by a sum of component meanings taking into account diagnostic problem specificity.

The algorithm of respective component estimation relies on an iterative alternate matched filtering and thresholding scheme [18].

3.5 Patch Domain Modeling

Local patches are block contexts of each or selected pixels to be decomposed according to multi-component sparse criteria. According to [19], an image $f \in \mathcal{R}^N$ of N pixels is processed by extracting patches $p(f_x)$ of size $\nu \times \nu$ around selected pixel position $x \in \{1, \ldots, X\}$, $X \leq N$.

$$\forall_{t \in \{-\nu/2+1, \ldots, \nu/2\}^2}, p_t(f_x) = f_{x+t} \tag{2}$$

Thus, the patch $p(f_x)$ is handled as a vector of size $n = \nu^2$. Next, linear modeling operator $\mathcal{P} : f \rightarrow \{p(f_x)\}_x$ extracts all the patches from an image.

Each patch is approximated M-sparse representation of local features as

$$\overset{\approx}{f}_x = \sum_{k=1}^{M} w_{j_k} \varphi_{j_k} \tag{3}$$

where each $w_{j_k} \in \mathbf{R}$ is sorted coefficient associated to the atom $\varphi_{j_k} \in \mathbf{R}^n$ from patch-adjusted dictionary $\mathcal{D}_x = \{\varphi_j\}$.

4 Experimental Conclusions

Presented methodology of image understanding was implemented and developed in several applications of computer-aided diagnosis, i.e. ischemic stroke diagnosis, screening mammography diagnosis or videobronchoscopy - based tumor mass detection [20–22].

Sparse, hierarchical and multiscale image representation occurred extremely useful for subtle or hidden signal extraction because its capability for signal energy packing with preserved localization across scales and subbands. Selection of specific decomposition atoms adjusted to crucial image features allow target content modeling and extraction through sparse data representation.

Usefulness of sparse representation was confirmed by effective extraction of hidden diagnostic information in case of acute ischemia detection. Accurate early diagnosis of hyperacute ischemic stroke is critical due to limited timing of applicable thrombolytic therapy. However, clinical phenotype is today obligatory completed with neuroimaging. It should allow identification of patients with acute stroke and selection of suitable treatment. Computed tomography (CT) as an imaging method of first choice is used for efficient identification of patients with acute stroke. Consequently, it allows selection of suitable treatment, exclusion of intracerebral hemorrhage and determination of etiology as well as follow-up therapy and its possible complications.

A CT image of the brain in acute stroke patients is not self-evident. Reading of CT needs training and additional knowledge about the physical conditions of image contrast distribution with noise and artifacts-caused limitations [23]. Significant CT number instability masks very subtle hypodense changes within ischemic region making pathology detection extremely difficult for many cases of irreversible infarcts. Thus, a challenge for CAD applications is making hypodensity distribution more distinct to reveal the diagnostic content and improve accurate recognition of infarct signatures.

The sparse model of masked tissue density was used to extract subtle, diagnostically important hypodensic changes. Multiscale image processing method was designed according to semantic sparse representation paradigm:

- initial estimation of diagnostic content:
  - image conditioning with segmentation of stroke-susceptible regions of brain tissues – locally adaptive region growing and thresholding methods with smooth complement of segmented diagnostic areas,
  - target content estimation with adjusted tensor wavelet kernel – orthogonal filter bank defined by low pass filter $\tilde{h} = [1/4, 2/4, 1/4]$;
- sparse data representation for subtle hypodensity signs extraction – selected curvelet atoms through nonlinear approximation with "waveshrink";

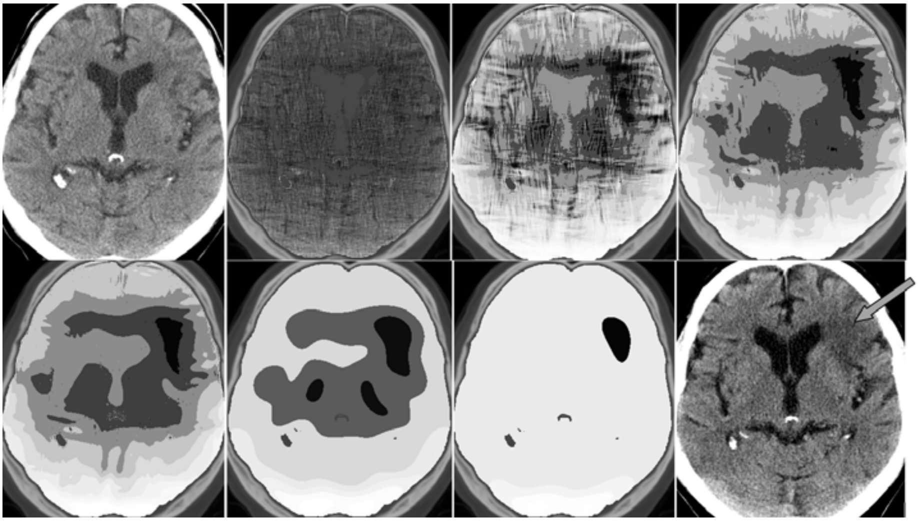


Fig. 1. Disease signature extraction based on sparse image representation. The following images (left to right, top-down) are: –acute CT scan, –estimated density distribution with reduced noise, –five successive hypodensity visual expressions from sparse representation consisting of: 27% of nonzero coefficients, 6%, 2.5%, 1.2%, 0.06% and 0.005% of nonzero coefficients, –follow-up CT confirming ischemic stroke (by indicated area of hypodensity).

– visual hypodensity expression - display arrangement of processed regions and source scans with greylevel quantization and contrast enhancement according to observer suggestions and semantic content models;
– sparse data classification in diagnostic categories – automatic recognition of the regions of ischemic brain tissue - texture features extracted in spatial and sparse multiscale domain, selected and classified with SVM.

More details of paradigm implementation was given in [20, 25]. Exemplary results of hypodensity extraction dependent on a representation sparsity were presented in Fig. 1.

Achieved results are promising and the potential of semantic sparse data representation for diagnostic content modelling seems to be almost unlimited. But clue optimization problems are: – new atoms of multiscale, local, flexible image approximants adjusted to representative pathology patterns, – more effective schemes of target content estimation, – investigation of pursuit algorithms of best basis matching based on diagnostic knowledge criteria. Comparative study to other automatic image understanding concepts including theoretical analysis and experimental verification are planned.

References

1. Garland, L.H.: Studies on the accuracy of diagnostic procedures. AJR 82, 25–38 (1959)
2. Borgstede, J.P., Lewis, R.S., Bhargavan, M., Sunshine, J.H.: RADPEER quality assurance program: a multifacility study of interpretive disagreement rates. J. Am. Coll. Radiol. 1, 59–65 (2004)
3. Berlin, L.: Accuracy of diagnostic procedures: has it improved over the past five decades? AJR 188, 1173–1178 (2007)
4. Renfrew, D.L., Franken, E.A., Berbaum, K.S., Weigelt, F.H., Abu-Yousef, M.M.: Error in radiology: classification and lessons in 182 cases presented at a problem case conference. Radiology 183, 145–150 (1992)
5. Oestmann, J.W., Greene, R., Kushner, D.C., Gourgouin, P.M., Linetsky, L., Llewellyn, H.J.: Lung lesions: correlation between viewing time and detection. Radiology 166, 451–453 (1988)
6. Kan, L., Olivotto, I.A., Burhenne, L.J.W., Sickles, E.A., Coldman, A.J.: Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program. Radiology 215, 563–567 (2000)
7. Loy, C.T., Irwig, L.: Accuracy of diagnostic tests read with and without clinical information: a systematic review. JAMA 292, 1602–1608 (2004)
8. Smith, M.J.: Error and variation in diagnostic radiology. Springfield, IL: C C Thomas 4 71,73,74,144–169 (1967)
9. Greene, R.E.: Missed lung nodules: lost opportunities for cancer cure. Radiology 182, 8–9 (1992)
10. Tadeusiewicz, R.: Automatic understanding of signals. In: Intelligent Information Processing and Web Mining, Proc of the International Intelligent Information Systems, IIPWM 2004 Conference, Zakopane 2004, Springer, Heidelberg (2004)

11. Ogiela, M., Tadeusiewicz, R.: Modern computaptional intelligence methods for the interpretation of medical images. Studies in Computational Intelligence, vol. 84. Springer, Heidelberg (2008)
12. Tadeusiewicz, R., Ogiela, M.: Automatic image understanding. A new paradigm for intelligent medical image analysis. Bio-Algorithms and Med-Systems Journal edited by Medical College - Jagiellonian University 2(3), 5–11 (2006)
13. Ogiela, M., Tadeusiewicz, R.: Nonlinear processing and semantic content analysis in medical imaging - a cognitive approach. IEEE Trans. Instrum. Meas. 54(6), 2149–2155 (2005)
14. Blum, A., Langley, P.: Selection of relevant features and examples in machine learning. Artificial Intelligence 97, 245–271 (1997)
15. Podsiadly-Marczykowska, T., Guzik, A.: Mammography ontology, model structure, definitions and conception instances. Bio-Algorithms and Med Systems 1(1), 247–252 (2005)
16. Ancona, N., Maglietta, R., Stella, E.: Data representation in kernel based learning machines. In: Machine Learning and Applications, Proceedings, pp. 129 - 136 (2004)
17. Welland, G.V. (ed.): Beyond Wavelets. Studies in Computational Mathematics, vol. 10. Academic Press, London (2003)
18. Bobin, J., Starck, J.-L., Fadili, J.M., Moudden, Y., Donoho, D.L.: Morphological component analysis: an adaptive thresholding strategy. IEEE Trans. Im. Proc. 16(11) (2007)
19. Peyre, G.: Sparse modeling of textures. J. Math. Im. Vis. 34(1), 17–31 (2009)
20. Przelaskowski, A., Sklinda, K., Ostrek, G., Józwiak, R., Walecki, J.: Computer aided diagnosis in hyper-acute ischemic stroke. In: Walecki, J. (ed.) Progress in Neuroradiology 2009, pp. 69–78. International Scientific Literature, Inc., New York (2009)
21. Przelaskowski, A., Podsiadly-Marczykowska, T., Wroblewska, A., Boninski, P., Bargiel, P.: Computer-aided interpretation of medical images: mammography case study. Machine Graphics & Vision 16(3/4), 347–376 (2007)
22. Przelaskowski, A., Józwiak, R., Zieliñski, T., Duplaga, M.: Endobronchial tumor mass indication in videobronchoscopy - block based analysis. In: Proceedings of VISAPP 2010 (2010) (in press)
23. von Kummer, R.: The impact of CT on acute stroke treatment. In: Lyden, P. (ed.) Thrombolytic Therapy for Stroke. Humana Press, Totowa (2005)
24. DeVore, R.A.: Nonlinear approximation. Acta Numerica 7, 51–150 (1998)
25. Przelaskowski, A., Ostrek, G., Sklinda, K., Walecki, J., Józwiak, R.: Stroke slicer for CT-based automatic detection of acute ischemia. In: Advances in Intelligent and Soft Computing. Computer Recognition Systems 3, vol. 57, pp. 447–454. Springer, Heidelberg (2009)

Semantic Interpretation of Heart Vessel Structures Based on Graph Grammars

Mirosław Trzupek

AGH University of Science and Technology, Institute of Automatics
Al. Mickiewicza 30, PL-30-059 Kraków, Poland
mtrzupek@agh.edu.pl
http://home.agh.edu.pl/~mtrzupek/

Abstract. In this paper discussed are the manners of applying methods of computational intelligence to the development of a syntactic semantic description of spatial visualisations of the heart's coronary vessels. The IT methodology described in this work makes it possible to find, for each biological 3D structure considered, its semantic description whose elements make reference to the medical significance of the entire structure described, while at the same time refrain from the formal differences of individual visualisations.

Keywords: Medical image understanding, syntactic pattern analysis, artificial intelligence, Computer-Aided Diagnosis.

1 Introduction

Despite the huge progress in the equipment that allows obtaining a great variety of coronary vessel visualisations, and especially the latest achievements that allow their 3D visualisation, the progress in the automatic analysis of these images is highly limited [3, 4, 7, 12, 13]. Most research is still made manually by the operator or diagnostician and unfortunately, no computer-assisted systems offering automatic interpretation comparable to that conducted by the specialist's keen eye. For that reason, the article in its further part describes the manner of developing linguistic semantic representation of spatial structure of coronary vessels showing certain pathological changes and their analysis with the use of cognitive methods of image understanding [8, 11]. Thanks to such a description, it will later be possible to support interpretative tasks and define the semantics of actual pathologies. The technical solutions proposed do not aim at the elimination of the physician from the diagnostic process, on the contrary, they should provide the physician with additional information to provide him or her with higher efficiency in discrimination between obvious cases from ones that may require a more thorough consideration. The problem embarked upon in this work may be considered significant due to the frequency of practical use of identification of location of stenoses in coronary vessels in ischemia, being one of the most burning medical problems of the early 21st century [2]. A system for understanding 3D medical images may also facilitate interpretation of especially complex and atypical cases, when even the physician's knowledge combined with experience may be insufficient for diagnosing and defining treatment strategy.

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 81–88, 2010.

2 Modelling and the Meaning Analysis of Coronary Vessel Reconstructions

To allow the introduction of linguistic formalisms in the form of graph-based grammar, several visualisations sequences were performed during diagnostic examinations of the heart with a helical CT scanner. Such visualisations present in a very clear manner all morphologic changes of individual sections of arteries in any plane. The structures made visible in this manner will be described with the use of graph-based grammars - constituents of the graph-based language defining their proper, spatial topology [8–11]. In order to analyze a 3D reconstruction, it becomes necessary to select the appropriate projection showing lesions in vessels in a way that enables them to be analysed on a plane. In the clinical practice, this operation is done manually by the operator. In our research we have attempted to automate the procedure of finding such a projection by using selected geometric transformations during image processing. Using the fact that the spatial layout of an object can be determined by projecting it onto the axes of the Carthesian coordinate system, values of horizontal Feret diameters, which are a measure of the horizontal extent of the diagnosed coronary artery tree, are calculated for every subsequent animation frame during the image rotation. The projection for which the horizontal Feret diameter is the greatest is selected for further analyses, as this visualisation shows both the right and the left coronary artery in the most convenient take (Fig. 1).

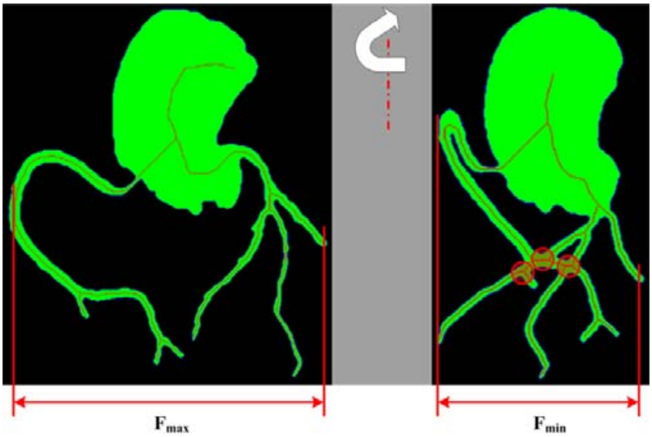


Fig. 1. Projections for which the horizontal Feret diameter are the greatest (A) and the smallest (B)

To enable a linguistic representation of the spatial reconstructions studied, the coronary vessels shown in them had been subjected to the operation of thinning, referred to as skeletonising. The skeleton can then be subjected to the operation of labelling, which determines the start and end points of main and surrounding branches of coronary arteries in it. These points will constitute the peaks of a graph modelling the spatial structure of the coronary vessels of the heart. The next step is labelling them by giving each located informative point the appropriate label from the set of peak labels which unambiguously identify individual coronary arteries forming parts of the structure analysed. In the case of terminal points (leaves of a graph modelling the coronary vascularisation), the set of peak labels comprises abbreviated names of arteries found in coronary vascularisation and have been defined as follows:

a) for the left coronary artery: LCA - left coronary artery, LAD - anterior interventricular branch (left anterior descending), CX - circumflex branch, L - lateral branch, LM - left marginal branch
b) for the right coronary artery: RCA - right coronary artery, A - atrial branch, RM - right marginal branch, PI - posterior interventricular branch, RP - right posterolateral branch.

If a given informative point is a branching point, then the peak will be labelled with the concatenation of names of the peak labels of arteries which begin at this point. This way, all initial and final points of coronary vessels as well as all points where main vessels branch or change into lower level vessels have been determined and labelled as appropriate. After this operation, the coronary vascularisation tree is divided into sections which constitute the edges of a graph modelling the examined coronary arteries. This makes it possible to formulate a description in the form of edge labels which determine the mutual spatial relations between the primary components, i.e. between subsequent arteries shown in the analysed image. These labels have been identified according to the following system. Mutual spatial relations that may occur between elements of the vascular structure (represented by a graph) are described by the set of edge labels. The elements of this set have been defined by introducing the appropriate spatial relations: vertical, defined by the set of labels $\{\alpha, \beta, \ldots, \mu\}$ and horizontal, defined by the set of labels $\{1, 2, \ldots, 24\}$ on a hypothetical sphere surrounding the heart muscle. These labels designate individual final intervals, each of which has the angular spread of $15°$. Then, depending on the location, terminal edge labels are assigned to all branches identified by the beginnings and ends of the appropriate sections of coronary arteries. The presented methodology draws upon the method of determining the location of a point on the surface of our planet in the system of geographic coordinates, where a similar cartographic projection is used to make topographic maps. This representation of mutual spatial relationships between the analysed arteries yields a convenient access to and a unanimous description of all elements of the vascular structure. At subsequent analysis stages, this description will be correctly formalised using ETPL(k) graph grammars [9, 10], supporting the search for stenoses in the lumen of arteries forming parts of the coronary vascularisation.

As the structure of coronary vascularisation may be characterised by three different types of artery distribution over the heart surface, in the following part we will propose a grammar for the left dominance artery distribution. The left dominance artery distribution is present on average 10-14% of cases [2, 6]. Before we define the representation of the analysed image in the form of IE graphs, we have to introduce the following order relationship in the set of Γ edge labels: $1 \leq 2 \leq 3 \leq ... \leq 24$ and $\alpha \leq \beta \leq \gamma \leq \ldots \leq \mu$. This way, we index all peaks according to the $\leq$ relationship in the set of edge labels which connect the main peak marked 1 to the adjacent peaks and we index in the ascending order ($i = 2, 3, \ldots, n$). This gives us IE graphs for the right and the left coronary arteries (Fig. 2).

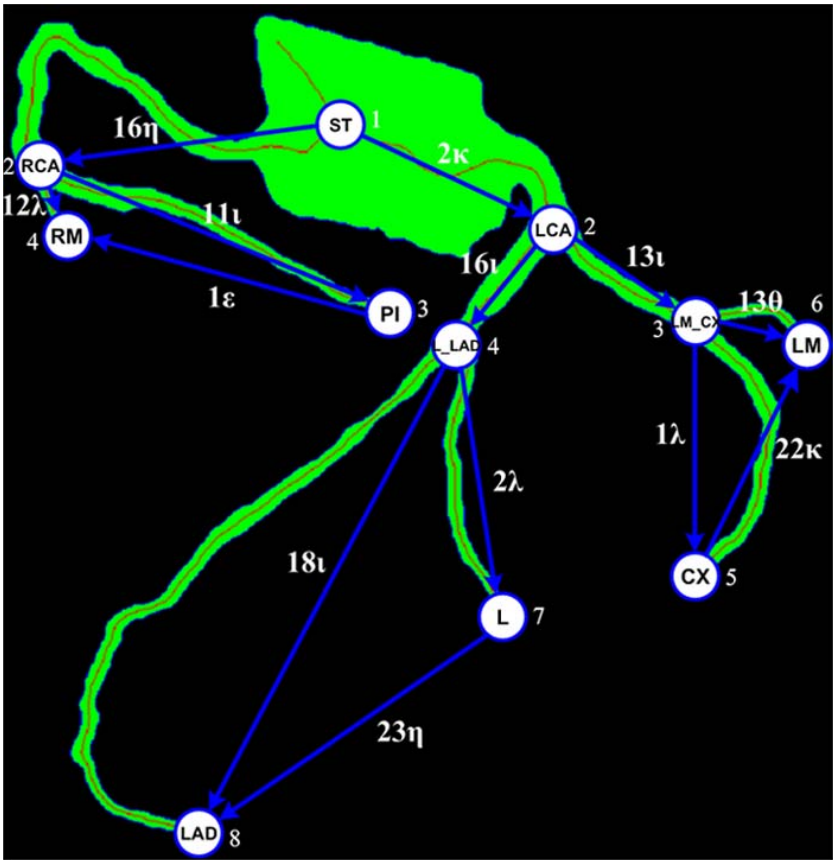


Fig. 2. The representation of the left and the right coronary artery using IE graphs

When graphs shown in Fig. 2 are represented by their characteristic descriptions, they look as presented in Table 1 and Table 2.

The graph structure created in this way will form elements of a graph language defining the spatial topology of the heart muscle vascularisation including its

Table 1. Description for the right coronary artery

ST_1	RCA_2	PI_3	RM_4
1	2	1	–
16η	11ι 12λ	1ϵ	–
2	3 4	4	–

Table 2. Description for the left coronary artery

ST_1	LCA_2	LM_CX_3	L_LAD_4	CX_5	LM_6	L_7	LAD_8
1	2	2	2	1	–	1	–
2κ	13ι 16ι	1λ 13θ	2λ 18ι	22κ	–	23η	–
2	3 4	5 6	7 8	6	–	8	–

possible morphological changes. For IE graphs defined as above, in order to locate the place where stenoses occur, the graph grammar may take the following form.

a) for the right coronary artery: $G_R = (\Sigma, \Delta, \Gamma, P, Z)$
$\Sigma = \{ST, RCA, PI, RM, C\_Right\}$
$\Delta = \{ST, RCA, PI, RM\}$
$\Gamma = \{16\eta, 11\iota, 12\lambda, 1\epsilon\}$
The start graph Z and the set of productions P shown in Fig. 3.

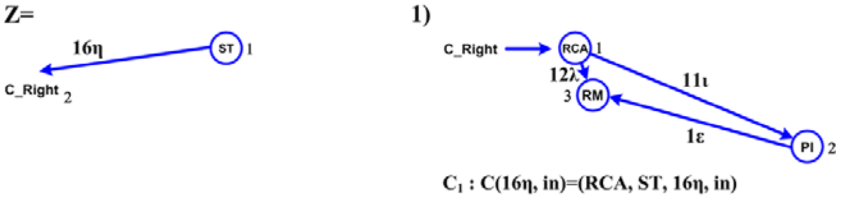
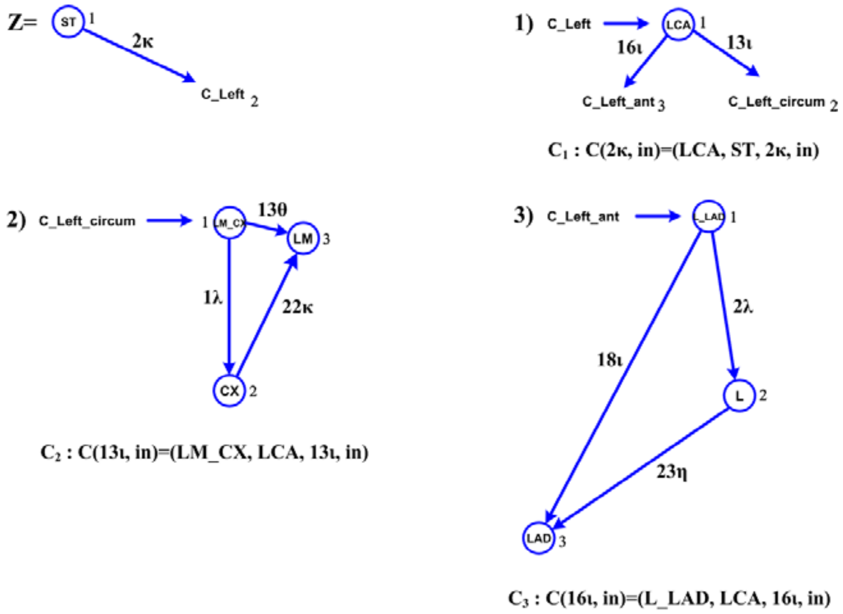


Fig. 3. Start graph Z and the set of productions for grammar G_R

b) for the left coronary artery: $G_L = (\Sigma, \Delta, \Gamma, P, Z)$
$\Sigma = \{ST, LCA, LM\_CX, L\_LAD, CX, LM, L, LAD, C\_Left, C\_Left\_ant, C\_Left\_circum\}$
$\Delta = \{ST, LCA, LM\_CX, L\_LAD, CX, LM, L, LAD\}$
$\Gamma = \{2\kappa, 13\iota, 16\iota, 13\theta, 22\kappa, 1\lambda, 2\lambda, 18\iota, 23\eta\}$
The start graph Z and the set of productions P shown in Fig. 4.

This way, we have defined a mechanism in the form of ETPL(k) graph grammars which create a certain linguistic representation of each analyzed image in the form of an IE graph [9, 10]. The set of all representations of images generated by this grammar is treated as a certain language. Consequently, we can built a syntax analyzer based on the proposed graph grammar which will recognize elements of this language. Theoretical aspects of conducting the syntax analysis



C_1 : C(2κ, in)=(LCA, ST, 2κ, in)

C_2 : C(13ι, in)=(LM_CX, LCA, 13ι, in)

C_3 : C(16ι, in)=(L_LAD, LCA, 16ι, in)

Fig. 4. Start graph Z and the set of productions for grammar G_L

and constructing parsers for ETPL(k) grammars are described in [9, 10]. It is worth remembering at this point that for the graph grammar class considered here, there are deterministic automatons (parsers) which perform syntax analysis in a multinomial time, characterised by the complexity of $O(n^2)$.

3 Results

The proper element of the syntactic and semantic analysis is parsing that, in the case of structural analysis of the representations developed, provides automatically (thanks to the use of semantic actions and procedures) full information defining the spatial topology of the graph in question, describing the coronary vessels together with the individual constituents. Subsequently, such information is used for the structural and semantic analysis of individual sections of the arteries. The recognition of stenoses, including the determination of their locations, lumens of the artery, and the types (concentric or eccentric), was conducted in such a way that while reasoning out the grammar for the graph representation of the coronary vascularisation, particular edges of the graph determined the actual beginnings and ends of particular sections of coronary arteries. During the grammar reasoning and the course of the transform of embedding graph representations on the actual images, the corresponding sections of arteries were analysed with regard to the presence of potential stenoses in them. The method of this analysis also consisted in applying a context-free sequential grammar to detect stenoses in 2D coronarography images. Such a grammar has been defined in publication [9–11]. Applying a context-free sequential grammar to analyse particular sections of arteries in the obtained spatial reconstructions turned out to be quite effective, as it allowed the unanimous location of the lesion present together with defining their size and type. On the image data tested, the efficiency of recognition (locating and defining the type of stenosis) amounted to 85%. In order to estimate the correct value of the narrowing degree, obtained with the use of grammars semantic actions, several comparative values from syngo Vessel View [14] clinical application were used. Such system is often used in clinical practice with the SOMATOM Sensation Cardiac 64 [14], which was equipped with vessel segmentation routines allowing for measurement stenosis quantification.

4 Conclusion

The results achieved show that, thanks to the application of AI graph-based linguistic formalisms, allowing semantic interpretation of spatial reconstruction of coronary vessels, it is possible to analyse the correct morphology of individual sections of arteries and their branches. Furthermore, applying methods of computational intelligence to the development of a syntactic semantic description of spatial visualisations of the heart's coronary vessels may be used for both smart ordering of images while archiving them and for their semantic searches. This is a significant innovation among the existing approaches to the computer-assisted medical diagnostics and broadens greatly the scope of use of AI methodologies towards the computer-assisted methods of understanding diagnostic images and the definition of medical significance of pathologies present therein.

Acknowledgement

This work has been supported by the AGH University of Science and Technology under Grant No. 10.10.120.783.

References

1. Bankman, I. (ed.): Handbook of Medical Imaging: Processing and Analysis. Academic Press, London (2002)
2. Faergeman, O.: Coronary Artery Disease. Elsevier, Amsterdam (2003)
3. Higgins, W.E., Karwoski, R.A., Ritman, E.L., Spyra, W.J.T.: System for analyzing true three-dimensional angiograms. IEEE Trans. Med. Imag. 15, 377–385 (1996)
4. Hoffman, K.R., Sen, A., Li, L.: A system for determination of 3D vessel tree centerlines from biplane images. The International Journal of Cardiac Imaging 16, 315–330 (2000)
5. Meyer-Baese, A.: Pattern Recognition in Medical Imaging. Elsevier-Academic Press, San-Diego (2003)
6. Khan, M.G.: Heart Disease Diagnosis and Therapy. Williams & Wilkins, Baltimore (1996)

7. Lewandowski, P., Tomczyk, A., Szczepaniak, P.S.: Visualization of 3-D Objects in Medicine - Selected Technical Aspects for Physicians. Journal of Medical Informatics and Technologies 11, 59–67 (2007)
8. Ogiela, M.R., Tadeusiewicz, R.: Modern Computational Intelligence Methods for the Interpretation of Medical Images. Springer, Heidelberg (2008)
9. Skomorowski, M.: A Syntactic-Statistical Approach to Recognition of Distorted Patterns. Jagiellonian University, Kraków (2000)
10. Tadeusiewicz, R., Flasiński, M.: Pattern Recognition. PWN, Warsaw (1991)
11. Tadeusiewicz, R., Ogiela, M.R.: Medical Image Understanding Technology. Springer, Heidelberg (2004)
12. Wild, P.S., Zotz, R.J.: Fragment reconstruction of coronary arteries by transesophageal echocardiography - A method for visualizing coronary arteries with ultrasound. Circulation 105, 1579–1584 (2002)
13. Wróbel, K., Porwik, E., Porwik, P.: Three dimensional image projections and its measurement using the vrml technique. Journal of Medical Informatics and Technologies 11, 123–133 (2007)
14. Get the Entire Picture, SOMATOM Sensation Cardiac 64 Brochure. Siemens medical (2004)

Interpretation of Images and Their Sequences Using Potential Active Contour Method

Stanisław Walczak, Arkadiusz Tomczyk, and Piotr S. Szczepaniak

Institute of Information Technology, Technical University of Łódź
ul. Wólczańska 215, 90-924 Łódź, Poland
walczak@ics.p.lodz.pl, tomczyk@ics.p.lodz.pl

Abstract. The aim of this paper is to present three approaches to cardiac ventricle segmentation, which apply the potential active contour method. Two of these approaches use three-dimensional, and one of them - four-dimensional representation of data. The approaches presented simulates expert's behaviour. They aim at image segmentation of cardiac ventricles performed at all slices simultaneously, thanks to which every slice can be analysed in the context of knowledge about other slices. The medical image understanding method is not fully automatic, however in comparison to manual segmentation performed by an expert, it saves much time, which may be of vital importance for patient's health e.g. in pulmonary embolism diagnosis.

1 Introduction

The aim of this paper is to present and compare three semi-automatic methods of cardiac ventricle image segmentation, which may support pulmonary embolism diagnosis. Pulmonary embolism is a potentially lethal condition, which consists in a blockage of the lung artery. Consequently, the heart has difficulty in pumping blood to the lungs, which brings on high blood pressure. This, in turn, results in changes of ventricles' shape. The ventricles are enlarged and the interventricular septum becomes deformed. Such changes can be detected by analysing a series of ventricle images. The methods discussed are applied to four-dimensional images, which present the cardiac cycle. The methods currently used have proved to be very time-consuming, since they require manual detection of ventricle shape on each slice. Automation of this process can save much time, which is of vital importance for the patients.

Expertise shows that it is hard to determine the shape of ventricles on the basis of one image, separated from the others. Therefore, the methods presented attempt at retrieving heart contour by the analysis of image sequences (time sequences, space sequences, or both). Such an approach simulates expert's behaviour. Partial automation shortens the time needed for heart shape detection. The segmentation is performed by the potential active contour method. Its main advantage is the ability to detect cylindric, rounded shapes, which is particularly important in medical image analysis. In order to achieve a similar effect with

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 89–96, 2010.

other active contour methods one would have to define additional components of energy function responsible for smoothness of the contour. Another advantage of the method presented is the ability to adapt itself easily to both three- and four-dimensional data. Examples of the results are shown in Fig. 1.

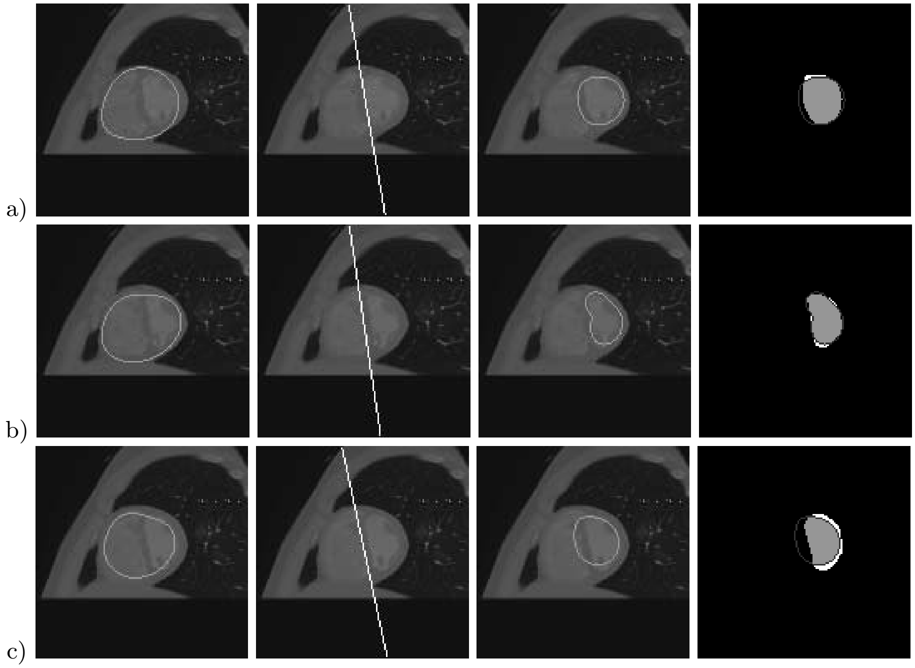


Fig. 1. Exemplary heart slices, respectively: contour of both heart ventricles, line indicating interventricular septum and contour of left ventricle (a) - contour detected by 3D space sequence analysis, (b) - contour detected by 3D time sequence analysis, (c) - contour detected by 4D approach

2 Potential Active Contours

Potential active contour method is one of the methods used for automatic image analysis. Supported by expert knowledge, it can be used for context image segmentation. The method was described in detail in [9]. Contrary to the typical, two-dimensional active contour methods, such as snakes that require additional modification to be used in higher-dimensional space, potential active contours make it possible to define contour model in space R^n (potential hypercontour model). The contour is defined as a set of points for which the values of fields generated by two types of sources are equal:

$$\left\{ \mathbf{x} \in \mathsf{R}^n : S^o(\mathbf{x}) = S^b(\mathbf{x}) \right\} \tag{1}$$

where:

$$S^o(\mathbf{x}) = \sum_{i=0}^{N^o} P_{\Psi_i^o,\mu_i^o}(\rho(\mathbf{x}_i^o, \mathbf{x})) \tag{2}$$

$$S^b(\mathbf{x}) = \sum_{i=0}^{N^b} P_{\Psi_i^b,\mu_i^b}(\rho(\mathbf{x}_i^b, \mathbf{x})) \tag{3}$$

In the above equations ρ denotes a metric in R^n and P is a function defining characteristic of the potential field. In the methods presented, it has been assumed that:

$$P_{\Psi,\mu}(d) = \frac{\Psi}{1 + \mu d^2} \tag{4}$$

Sources N^o are called object sources, whereas sources N^b denote background sources.

For contour evaluation, it is necessary to define objective function the value of which will be optimised in subsequent iterations. The objective function is usually a sum of several scalar energy functions. Their form depends closely on the method employed and its application, which will be described in detail below. The optimisation is performed using the simulated annealing algorithm, the fundamental advantage of which is that it requires defining only the objective function, without the necessity to define its derivatives. These derivatives are often very hard to define or it is impossible to calculate them analytically. In potential active contour method, the contour is determined by potential sources, thus the location of potential sources and their values are optimised: $\mathbf{x}_i^o$, Ψ_i^o, μ_i^o for $i = 1, \ldots, N^o$ and $\mathbf{x}_i^b$, Ψ_i^b, μ_i^b for $i = 1, \ldots, N^b$.

3 Application

In the approach presented, potential active contours are used for both three- and four-dimensional segmentation. In the former case, either all space sequences (slices) connected to a given phase of the heart cycle, or all time sequences (frames) connected to a given slice are analysed parallelly. In the latter case, the segmentation is performed parallelly on all images i.e. on both time and space sequences. In all cases discussed, ventricle shape detection is based on knowledge gained from many slices, none image is analysed in separation from the other ones.

Irrespective of the kind of data analysed, the process of cardiac ventricle detection consists of several stages, which simulates actions performed by an expert. At the first stage, the general shape of both ventricles is detected. Then, the shape is made more specific. The next step is to detect the interventricular septum ([11,10]). Having found that information the left ventricle can be found. Further it is assumed that:

- $I \subset \mathsf{R}^n$ – represents the set of all image elements
- $C_{in} = \{\mathbf{x} \in I : S^o(\mathbf{x}) \geq S^b(\mathbf{x})\}$ – represents the subset of all image elements situated inside contour C

- $C_{out} = \left\{\mathbf{x} \in I : S^o(\mathbf{x}) < S^b(\mathbf{x})\right\}$ – the subset of all image elements situated outside contour C

Additionally, it is assumed that $S^b(\mathbf{x}) > S^o(\mathbf{x})$ for $\mathbf{x} \notin I$. Thus, the elements situated outside the image are considered as the background. The method presented is not fully automatic as it requires indicating the value of threshold T above which the heart is clearly visible in the images (the heart is brighter than the surrounding tissues thanks to the contrast the patient was injected during the examination). The value of T is indicated only once for a given set of images, which is much faster and easier than manual contour drawing.

3.1 Three-Dimensional Approach with Space Sequences Analysis

This method was described in [5]. It is similar to the method which applies time sequences, discussed below.

3.2 Three-Dimensional Approach with Time Sequences Analysis

With T value determined both ventricles can be detected. The first step is rough segmentation, performed by a minimal set of potential sources, i.e. one image source and one background source. The contour detected is evaluated by means of the following energy functions:

$$E(C) = E_{white}(C) + E_{black}(C) \tag{5}$$

where:

$$E_{white}(C) = \frac{\sum_{\mathbf{x} \in C_{in}} f_{white}(\mathbf{x})}{1 + |C_{in}|} \tag{6}$$

$$E_{black}(C) = \frac{\sum_{\mathbf{x} \in C_{out}} f_{black}(\mathbf{x})}{1 + |C_{out}|} \tag{7}$$

and:

$$f_{white}(\mathbf{x}) = \min_{\mathbf{y} \in I \wedge f_T(\mathbf{y}) = 1} (\rho(\mathbf{x}, \mathbf{y})) \tag{8}$$

$$f_{black}(\mathbf{x}) = \min_{\mathbf{y} \in I \wedge f_T(\mathbf{y}) = 0} (\rho(\mathbf{x}, \mathbf{y})) \tag{9}$$

Function $f_T(\mathbf{y})$ describes the image after thresholding:

$$f_T(x) = \begin{cases} 1 \text{ , } L(\mathbf{x}) > T \\ 0 \text{ , } L(\mathbf{x}) \leq T \end{cases} \tag{10}$$

where $L(\mathbf{x})$ is the brightness of an image element $\mathbf{x} \in I$. The next step, is a more precise retrieval of both ventricles. For this purpose, a configuration with more sources is used, which enables the contour to take more complicated shapes. Each source from the initial configuration is replaced by several new sources (a group of sources) situated in the same place as the original source. The sum of source potentials in the group equals the value of the original source potential. As a result, the initial contour in precise segmentation is identical to the final contour in rough segmentation. As the initial temperature in the simulated annealing algorithm decreases, contour optimisation is greedy. For contour evaluation, the following energy function is applied:

$$E(C) = E_{white}(C) + E_{black}(C) + E_{cohesion}(C) \tag{11}$$

where $E_{cohesion}(C)$ denotes energy member, which is responsible for cohesion of the contour. Its value is calculated for each slice separately and it equals the difference between the number of all pixels belonging to the inside part of the contour in a given slice and the number of pixels constituting a cohesive area of the contour.

Next, the interventricular septum is detected. This can be accomplished considering the following observation: the line must go through the previously detected heart contour and should also have a much lower value of brightness than the pixels forming the ventricles. This action is performed by means of a brute force algorithm. The algorithm searches through all segments that begin at the top edge and end at the bottom edge of the image and chooses the segment that goes through the largest number of dark pixels i.e. those with brightness level below T.

The last step is left ventricle segmentation. Analogously as with both ventricles, the segmentation is divided into two parts: the rough one - with a minimal set of potential sources, and the one with more sources. In both parts, an additional component is added to the evaluating function, which increases the value of energy for each pixel situated inside the contour but on the wrong side of the line representing interventricular septum.

The optimisation algorithm stops either after an initially set number of iterations or when the value of energy function does not change.

3.3 Four-Dimensional Approach

In the four-dimensional approach all images were analysed simultaneously. Energy function did not have to be redefined. However, the detection of interventricular septum was performed for each two-dimensional image independently.

4 Results

At each stage, different configurations of potential sources have been used. For initial shape detection, regardless of the number of dimensions, the minimal configuration $N^o = 1$, $N^b = 1$, has been used. For a more precise detection, the authors have applied a configuration which has resulted in detection of more complex shapes ($N^o = 4$, $N^b = 2$). Analogously, the left ventricle detection was initially performed by using a minimal configuration ($N^o = 1$, $N^b = 1$) and then

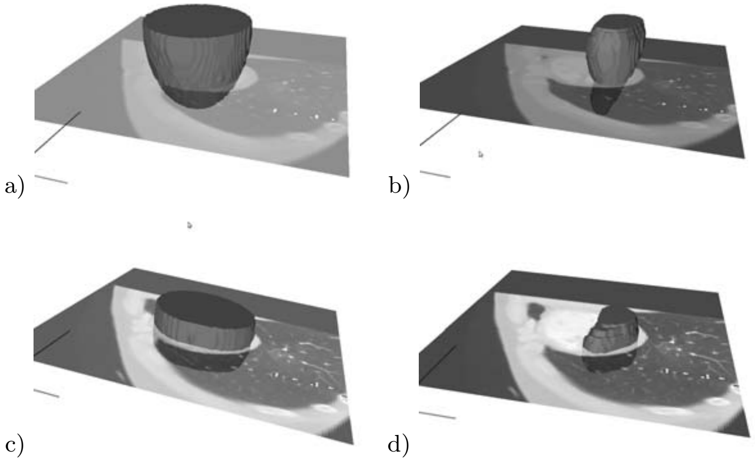


Fig. 2. 3D views with the corresponding slice in the background (a) - both ventricles, detected by 3D space sequence analysis, (b) - left ventricle, detected by 3D space sequence analysis, (c) - both ventricles, detected by 3D time sequence analysis, (d) - left ventricle, detected by 3D time sequence analysis

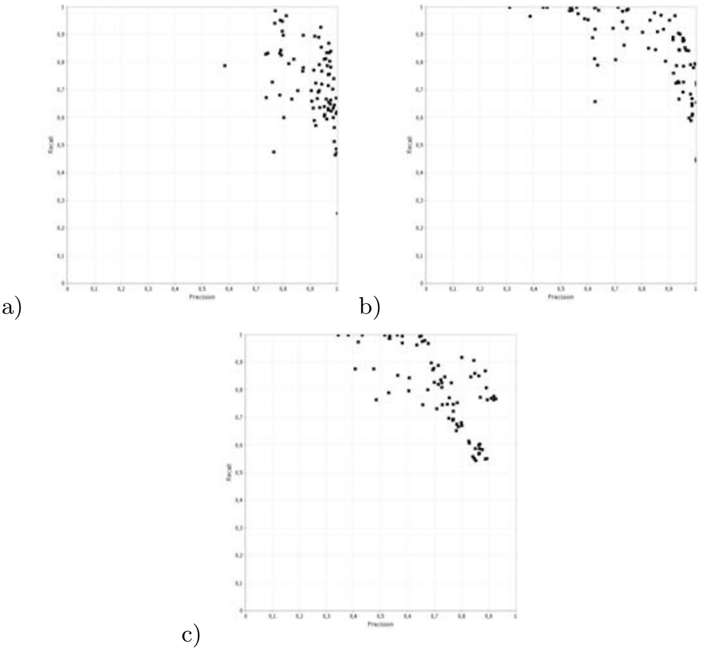


Fig. 3. Charts illustrating precision and recall values for the left ventricle slices segmentation: (a) - detection applying 3D space sequence analysis, (b) - detection applying 3D time sequence analysis, (c) - detection applying 4D analysis

more precisely, by using 4 object sources and 2 background sources respectively ($N^o = 4$, $N^b = 2$) for both three- and four-dimensional space.

The contours that were drawn by an expert are regarded as a model solution here. It has been assumed that they represent the desired division into pixels of the object (cardiac ventricles) and pixels of the background. Thus, it is possible, which was presented in [9], to evaluate the retrieved contours by precision and recall values where the contour is treated as binary classifier. Recall reflects the classifier's ability to retrieve all relevant objects, whereas precision provides information on how many objects retrieved by the classifier actually are relevant. The closer their score to 1.0, the better the results generated by the classifier. If the precision score is smaller than 1.0 , the retrieved contour is bigger than the model. The recall score smaller than 1.0 indicates that the classifier has omitted some of the pixels that have been indicated by an expert as relevant. The values of precision and recall for the same heart 4D image and three different segmentation approaches presented in this paper are depicted in Fig. 3.

5 Summary

The paper has presented three approaches to cardiac ventricles detection supported by potential active contour methods. All of them simulate behaviour of an expert. The three-dimensional method, which analyses space sequences, considers all images connected to the same moment of cardiac cycle simultaneously, whereas the method, which analyses time sequences, is based on various images of one slice of the heart recorded at various stages of the cardiac cycle. The four-dimensional method combines the two three-dimensional methods. Owing to this approach, each heart image may be segmented in the context of information gained from the other slices. The obtained results are promising, however, they reveal that context constituted by other slices and frames do not allow to gain contours that are close enough to the contours drawn by an expert. It suggests that other than pure visual information kind of knowledge, especially expert's knowledge and experience, should be incorporated into energy function. Those aspects are under further investigation.

Acknowledgement

This work has been supported by the Ministry of Science and Higher Education, Republic of Poland, under project number N 519 007 32/0978; decision no. 0978/T02/2007/32.

Authors would like to express their gratitude to Mr Cyprian Wolski, MD, from the Department of Radiology and Diagnostic Imaging of Barlicki University Hospital in Lodz for making heart images available and sharing his medical knowledge.

References

1. Gonzalez, R., Woods, R.: Digital Image Processing. Prentice-Hall Inc., New Jersey (2002)
2. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision, 321–331 (1988)
3. Kurzynski, M., Puchala, E., Wozniak, M., Zolnierek, A. (eds.): Computer Recognition Systems 2. Advances in Soft Computing, vol. 45. Springer, Heidelberg (2007)
4. Kurzynski, M., Wozniak, M. (eds.): Computer Recognition Systems 3. Advances in Soft Computing, vol. 57. Springer, Heidelberg (2009)
5. Walczak, S., Tomczyk, A., Szczepaniak, P.: Application of Contextual Active Contours to Heart Ventricle Shape Detection. Oprogramowanie Naukowo-Techniczne (2009)
6. Sonka, M., Hlavec, V., Boyle, R.: Image Processing, Analysis and Machine Vision. Chapman and Hall, Cambridge (1994)
7. Tadeusiewicz, R., Flasinski, M.: Rozpoznawanie obrazĂłw. Wydawnictwo Naukowe PWN, Warszawa (1991)
8. Tadeusiewicz, R., Ogiela, M.R.: Medical Image Understanding Technology, Studies in Fuzziness and Soft Computing, vol. 156. Springer, Heidelberg (2004)
9. Tomczyk, A.: Image segmentation using adaptive potential active contours. In: Kurzynski et al. [3], pp. 148–155
10. Tomczyk, A., Szczepaniak, P.S.: Segmentation of heart image sequences based on human way of recognition. In: Zhong, et al. [13], pp. 225–235
11. Tomczyk, A., Wolski, C., Szczepaniak, P.S., Rotkiewicz, A.: Analysis of changes in heart ventricle shape using contextual potential active contours. In: Kurzynski and Wozniak [4], pp. 397–405
12. Yezzi, A., Kichenassamy, S., Kumar, A., Olver, P., Tannenbaum, A.: A geometric snake model for segmentation of medical imagery. IEEE Transactions on Medical Imaging 16(2) (1997)
13. Zhong, N., Li, K., Lu, S., Chen, L. (eds.): BI 2009. LNCS, vol. 5819. Springer, Heidelberg (2009)

Inductive Learning Methods in the Simple Image Understanding System

Krzysztof Wójcik

Institute of Computer Science, Pedagogical University of Cracow
ul.Podchorążych 2, 30-084 Kraków, Poland
krzyw@ap.krakow.pl

Abstract. This article presents a proposition of using inductive learning methods in the task of creating the knowledge base for an image understanding system. With the help of the evolutionary algorithm, it is possible to synthesize an optimized system with the hierarchical structure of knowledge. The paper points to the key problem of the whole method - the creation of an effective algorithm of conceptual clustering. Some possible solutions are discussed.

1 Introduction

The most important characteristic of the image understanding idea is the extraction of the semantic content of the image (or more generally - the phenomenon). It is constructed by the automatic reasoning process, according to the knowledge about the images [6]. This scheme can essentially improve the automatic interpretation of the image. However, we can point to some weaknesses of this approach. The first one is the assumption that we have a useful model of the phenomena. The process of creating the model of a domain and defining the relationships between the concepts is often performed with the help of ontology. There are several tools which support the ontology building process but the starting point still depends on the knowledge and intuition of human being [5]. The second weakness is associated with a problem of knowledge acquisition - learning. We have a specific situation. The knowledge base applied to the image understanding system is relatively small, so the problems may be solved by simple, even primitive methods. What is more, typical image processing systems can easily collect a lot of data about the environment (we will assume that the environment is fully observable [5]). This encourages us to use inductive learning methods [3]. In order to use them in the concept creation process we should reduce our model to only necessary assumptions.

2 The Method Outline

The reduction of the assumptions and knowledge about the domain leads to the statement that the phenomenon consists of a kind of atomic, elementary parts, called primitive objects or instances. They correspond to primitive "concepts",

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 97–104, 2010.

called also types or classes. Suffice it to say that concepts are some kind of generalization of objects [1]. The objects are connected by primitive relationships. We will use the following notations:

$\mathbb{X}$ — space of all objects,
$\mathbf{X},\ \mathbf{X} \subset \mathbb{X}$ — set of objects,
$\mathbf{C_0}, \mathbf{C}$ — sets of concepts,
$\mathbf{D_0}, \mathbf{D}$ — sets of relationships.

At the beginning, we assume that $\mathbf{C_0}, \mathbf{D_0}$ represent primitive concepts and primitive relationships respectively. Additionally, $\mathbf{C} = \mathbf{C_0}, \mathbf{D} = \mathbf{D_0}$. These sets contain the whole initial knowledge. The main idea is that the general knowledge is included in the structure of concepts. So, the increasing of knowledge is possible by the creating new concepts, on the basis of visible objects. Let us assume, we can select the objects and recognize them as particular instances corresponding to the certain concepts (classification task). Let us also assume that it is possible to check all prior defined relations between all objects. Let us consider one of them, indicated by:

r_i, where: $r_i \in \mathbf{D},\ i \in \mathbf{I},\ \mathbf{I} = \{1, 2, \ldots, u\}$ is a set of indices of relations.

Let us presume, r_i has n arguments, so it may be satisfied by some n-tuple (ordered sequence of n elements) indicated by:

$t,\ t \in \mathbb{X}^n$ (nth Cartesian power of set $\mathbb{X}$)

Of course, this relationship may be satisfied by many other sequence, denoted by:

t_{ik}, where: $i \in \mathbf{I},\ k \in \mathbf{K},\ \mathbf{K} = \{1, 2, \ldots, m\}$ is a set of indices of tuples, which satisfy r_i relation (the first index of tuple t_{ik}, point to r_i relation).

Based on the selected tuples we will try to create a group of objects, that will allow us to construct the new concept. The objects should be in some way significant. We assume that it will happen if these objects will be connected by the relationships. We should also pay attention to a frequent "arrangement" of objects in the group, following the simple rule: *this phenomenon occurs many times, so it must be important.* Now we will be more precise about this. Let us define the group $\mathbf{G}$ as an ordered set of selected tuples:

$\mathbf{G} = \{t_{ik} : i \in \mathbf{I},\ k \in \mathbf{K}\}$

(we assume here that indexing of tuples allows us to identify their relations).

Let us transform the $\mathbf{G}$ set by the simple replacing of each object in each tuple by the label of object type. We obtain a set, denoted by $\mathbf{S}$, that describes an abstract[1] arrangement of relations in the group. The groups that have identical or similar[2] arrangement will be regarded as similar. Many such groups may appear in the sequence of input images. So, we should consider a set of similar groups:

$\mathbb{G} = \{\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_z\}$

The $\mathbb{G}$ set will be used in creation of the new concept. The next chapter presents a simple method of construction of the $\mathbb{G}$ set. Generally, this task should consist

[1] Depending on the types, not on the particular objects.

[2] Similarity may be defined here in several ways.

of a kind of conceptual clustering[3], in domain of defined above groups. After the clustering process we can obtain a set of groups, which are similar to each other, and according to this we can determine one, most characteristic group. We will use $\mathbf{G}'$ to denote this group, and $\mathbf{S}'$ to denote an arrangement of relations in $\mathbf{G}'$. The whole task of creation of $\mathbf{G}'$ set will be denoted by **FG**.

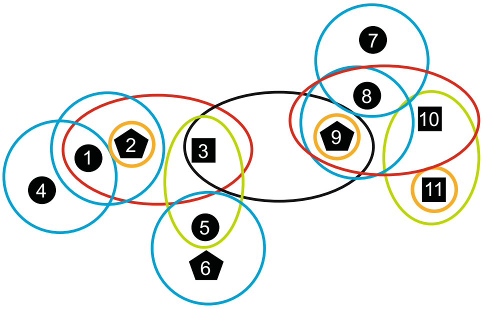


Fig. 1. Example of the group structure

Figure 1 shows several indexed objects $x_1, x_2, \ldots$ of three types (circle, pentagon, square). The few relationships are also defined (r_1 - red oval, r_2 - blue, r_3 - orange, r_4 - green, r_5 - black). We can specify several tuples:

$t_{31} = (x_2),\ t_{32} = (x_9),\ t_{33} = (x_{11})$ (the t_{31} denotes 1-st tuple of relation r_3)
$t_{21} = (x_1, x_4),\ t_{22} = (x_1, x_2),\ t_{23} = (x_8, x_9),\ t_{24} = (x_8, x_7)$
$t_{41} = (x_3, x_5),\ t_{42} = (x_{10}, x_{11}),$
$t_{11} = (x_1, x_2, x_3),\ t_{12} = (x_9, x_8, x_{10}).$

So, we can point to two groups "strongly" connected by the relations:

$\mathbf{G_1} = \{t_{31}, t_{21}, t_{22}, t_{11}\},\ \mathbf{G_2} = \{t_{32}, t_{24}, t_{23}, t_{12}\}$ then $\mathbb{G} = \{\mathbf{G}_1, \mathbf{G}_2\},\ \mathbf{G}' = \mathbf{G}_1$

Let us go back to the main problem. As a first approximation, we can say:

$$\textit{new concept} = \mathbf{S}'$$

The new concept is treated as an arrangement of relationships enriched by the information about the types of their arguments. To let us describe the new phenomena more completely, it should contain an additional information. We can define the object of a new concept as a combination of several sub-objects according to the structure of the group. The new objects may have new attributes (new properties vector). The objects should in some way inherit properties from sub-objects. It may be done for example by the simple copying, calculating a sum or average of given values. Generally, we can imagine the task of calculating

[3] We have a n-ary relations, that might be represented by the predicate $\mathbb{P}(t),\ t \in \mathbb{X}^n$. So, the solution of the clustering task may base on the usage of first-order representations and methods of inductive logic programming (ILP) [3].

the attributes as a combination of certain standard transformation. Let **FA** be a whole process of creating of new properties vector.

There is another significant question. The new objects are the combination of sub-objects, but may they be applied as arguments of old relationships? Or, should the old relationship be rather redefined in order to be applied to the new objects? It depends on how long we require the new objects to inherit a behavior of its parents. To be more precise, it depends on how the attributes are transformed (in the previous **FA** process) and what new interpretation they have. Let **FR** be a process of adapting or creating the new relationships.

Finally, we can consider the concept as a composition of elements:

$$\mathbf{S}', \mathbf{FG}, \mathbf{FA}, \mathbf{FR}$$

So, the creation of the new concept is performed by the combination of transformations **FG**, **FA** and **FR**. As a result we obtain the new concept, new relationships and new objects. We can join them into the $\mathbf{C}, \mathbf{D}, \mathbf{X}$ sets respectively. The whole process may be operated repetitively. This results in creating a hierarchical structure of concepts and hierarchical structure of new objects. We can say: in the structure of concepts the new general knowledge is included, while the structure of objects contains the knowledge about the particular scene.

But the fundamental question appears: is this knowledge really useful? The evaluation of the knowledge may be performed by the practical action. Let us imagine our system as an agent acting in the certain environment. According to the given scene and its own concept structure the agent can create the structure of objects. Then, we can compute the correlation between the existence of objects of certain concepts and the agent's ability to make correct decisions (or simple ability to survive). We assume that the agent be able to do the correct decisions has the most useful knowledge.

That suggests that to find the suitable concept structure we can use a "trial and error method" or a certain kind of evolutionary algorithm (a particular system is considered as an individual). However, the direct usage of such methods is problematical. We are fully aware that there are many possible kinds of transformation **FG**, **FA** and **FR** used repeatedly. Our solution is one point of the extremely large hypothesis space. In order to decrease the number of solutions, we can propose some heuristic approach. Let us assume that the selection of each of transformations is done with the help of a certain local criterion function. This function plays a role of a guide that helps searching through the decision tree. The next chapter provides some examples.

Let us consider the process **FG** of creation of groups again. This is a clustering process based on a prior defined performance function [5]. This function as well as the criterion functions mentioned earlier have a local character. The final, global evaluation of the agent system is performed by the genetic algorithm and depends on the agents capability to survive.

The situation depicted in the Figure 2 is an allegory that is associated with a simple example described in the next chapter.

Fig. 2. The indigenous inhabitants of the America and the Spanish conquistadors

3 The Simple Solution - The Example of Using the Method in the Image Analysis System

A general approach presented above is used in the simple image understanding system, designed mainly as a field of experiments. We will explain how the few simple concepts may be represented in the system. The example is trivial, but allows us to point to some important problems.

The first step refers to defining the initial sets $\mathbf{C_0}, \mathbf{D_0}$ (primitive concepts and relationships). As a system input is considered a sequence of pictures from CCD camera (we have used figures of animals drawn by a child). By using the segmentation method we can identify sets of pixels which create a simple raster figures. These figures may be easily transformed to vectors with properties like: *number of pixels, color, center coordinates, approximate shape.* This structure will be considered as a primitive object. Between such objects some primitive relationships are defined: binary - *neighboring*, and a few unary, for example: *having a small size, having a red color.* The next steps relate to the **FG**, **FA** and **FR** transformations.

1. The **FG** transformation, selection of the $\mathbb{G}$ and $\mathbf{G}'$ sets.

The relationships which hold between the objects in the groups generally refer to any, multi-argument predicates. In our simple case we will take into consideration only one- or two-argument relationships. A group **G** will be created by one main object $o_m \in \mathbf{X}$ and several objects which are connected with it by some relations. That group may be easily determined by checking all possible relationships of the o_m object. In this way we can create groups for all o_m objects. Having the set of groups, we can perform a simple clustering. In this connection we should draw attention to several universal criteria:

1. The count of similar groups. The groups which have identical arrangement of relationships are similar (see previous chapter), and may create a cluster. The count of groups in this cluster indicates how important it is. The arrangement of the groups will be called a ”pattern”.
2. The validity of objects that create the groups. After a few iterations of the concept creation process we obtain many combined objects. Each object may be evaluated directly basis on the count of elementary objects that were included.
3. The validity of the relationships between the objects of the group. In the simple way it can be expressed by the number of relationships that connect objects in the group. In addition we may respect the amount of one-argument relationships that are satisfied by the main o_m object.
4. The ”parent-child” restriction. Let us presume, we create a new group $\mathbf{G_n}$ based on the main object o_m. Let o_m (as a combined object) consists of ”parent” object o_p. Assume that the new group $\mathbf{G_n}$ includes also the o_p. Such a ”double” object as a redundant should be deleted or, if not, the evaluation of such a group should be decreased.
5. The similarity to old patterns. The similarity between groups on the analyzing scenes is of course a base of clustering. But the similarity between the new groups and groups that was already used to create the concepts is undesirable. It may result in creating a redundant concept.

According to the given rules we can make a sorted list of sets of groups. The best element that may be considered as the $\mathbb{G}$ set will be used to create the concept.
2. The **FA** transformation, definition of new properties vector.
The calculating of the attributes of the new objects must base on the properties of the component objects. Without going into details, we calculate the sum (*number of pixels*), average (*color, center coordinate*) and maximum (*shape*).
3. The **FR** transformation, process of adapting or creating the relationships.
We decide that the new concepts inherit all old relationships.

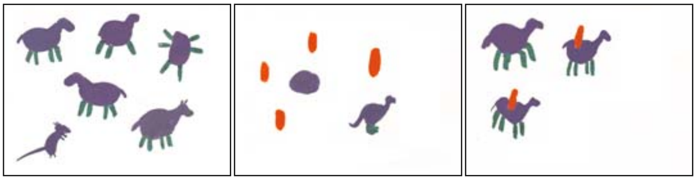
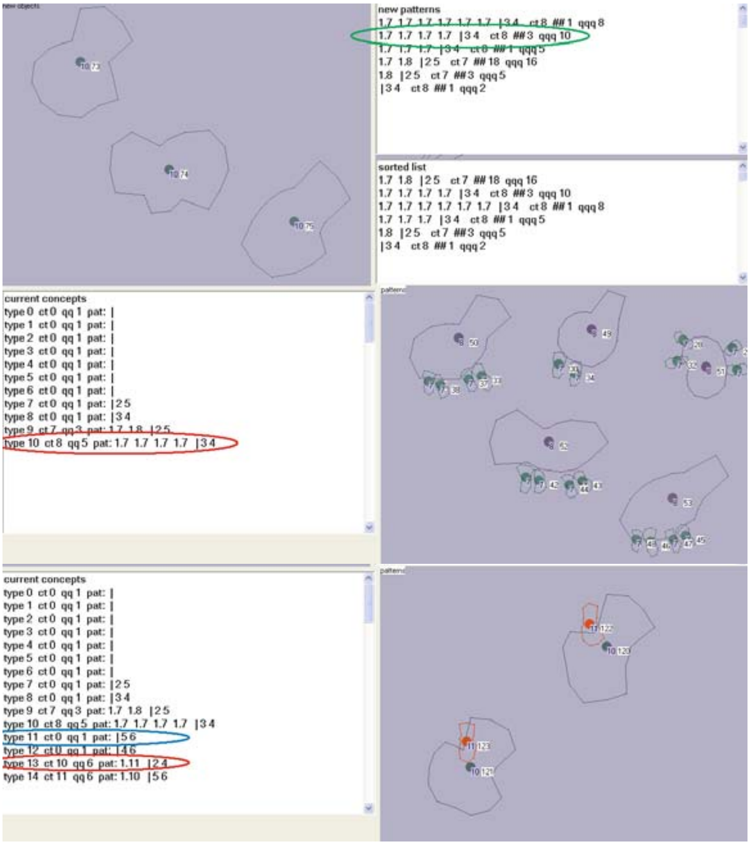


Fig. 3. The input sequence of pictures - the figures of animals

Figure 3 shows the sequence of analyzing pictures. The next one (Figure 4) presents processed images and frames showing the “state” of the system. The



```
new patterns
1.7 1.7 1.7 1.7 1.7 1.7 |3 4  ct 8 ## 1 qqq 8
1.7 1.7 1.7 1.7 |3 4  ct 8 ## 3 qqq 10
1.7 1.7 1.7 |3 4  ct 8 ## 1 qqq 5
1.7 1.8 |2 5  ct 7 ## 18 qqq 16
1.8 |2 5  ct 7 ## 3 qqq 5
|3 4  ct 8 ## 1 qqq 2
```

```
sorted list
1.7 1.8 |2 5  ct 7 ## 18 qqq 16
1.7 1.7 1.7 1.7 |3 4  ct 8 ## 3 qqq 10
1.7 1.7 1.7 1.7 1.7 1.7 |3 4  ct 8 ## 1 qqq 8
1.7 1.7 1.7 |3 4  ct 8 ## 1 qqq 5
1.8 |2 5  ct 7 ## 3 qqq 5
|3 4  ct 8 ## 1 qqq 2
```

```
current concepts
type 0  ct 0  qq 1  pat: |
type 1  ct 0  qq 1  pat: |
type 2  ct 0  qq 1  pat: |
type 3  ct 0  qq 1  pat: |
type 4  ct 0  qq 1  pat: |
type 5  ct 0  qq 1  pat: |
type 6  ct 0  qq 1  pat: |
type 7  ct 0  qq 1  pat: |2 5
type 8  ct 0  qq 1  pat: |3 4
type 9  ct 7  qq 3  pat: 1.7  1.8  |2 5
type 10  ct 8  qq 5  pat: 1.7  1.7  1.7  1.7  |3 4
```

```
current concepts
type 0  ct 0  qq 1  pat: |
type 1  ct 0  qq 1  pat: |
type 2  ct 0  qq 1  pat: |
type 3  ct 0  qq 1  pat: |
type 4  ct 0  qq 1  pat: |
type 5  ct 0  qq 1  pat: |
type 6  ct 0  qq 1  pat: |
type 7  ct 0  qq 1  pat: |2 5
type 8  ct 0  qq 1  pat: |3 4
type 9  ct 7  qq 3  pat: 1.7  1.8  |2 5
type 10  ct 8  qq 5  pat: 1.7  1.7  1.7  1.7  |3 4
type 11  ct 0  qq 1  pat: |5 6
type 12  ct 0  qq 1  pat: |4 6
type 13  ct 10  qq 6  pat: 1.11  |2 4
type 14  ct 11  qq 6  pat: 1.10  |5 6
```

Fig. 4. The interface of the image understanding system - the concept definitions

green circled item, top-right, is an example of a pattern (it is created from the first image on the Figure 3).

1.7 1.7 1.7 1.7 | 3 4 ct 8 ## 3 qqq 10[4]

The list of defined concepts is shown in the left-center frame. The red circled item:

type 10 ct 8 qq 5 pat: 1.7 1.7 1.7 1.7 | 3 4[5]

shows the definition of type 10, that represents something having big, blue body and four green legs. We can consider this as a concept "horse". The bootom

[4] 1.7 - relation 1 *neighboring*; holds between the main object and four others of type 7, 3 4 - properties of main object *having a blue color, having a big size*, ct 8 - type of the main object (index of concept on the list), ## 3 - amount of the group of objects that match the pattern, qqq 10 - value of performance function of pattern.

[5] type 10 - number of type, ct 8 - type of the main object, qq 5 - value of validity function of that type, pat: 1.7 1.7 1.7 1.7 | 3 4 - the pattern described earlier (objects of type 7 have properties 2,5; *having a green color, having small size*).

frames show the system state after analysis of second and third images presented in the Figure 3. We can identify two another concept:

type 11 ct 0 qq 1 pat: | 5 6

type 13 ct 10 qq 6 pat: 1.11 | 2 4

The first one we consider as a "rider", while the second as a "rider on the horse" (the object of type 11 "rider" is connected to the main object of type 10 "horse"). The creation of the last concept was possible when the concept "horse" was created first. We can say that the proper interpretation of a new phenomenon is possible when the suitable knowledge is available. Without them, it may be understood as a "dragon" with four legs and something red on the top. Let us go back to the illustration on the Figure 2. We can imagine what the young Indian and the old one (having wider knowledge) are thinking about. The proper understanding of the phenomenon allows them to make good, vital decision.

4 Conclusions

Despite the simplicity of the main idea, the attempt to make an implementation shows how many problems appear here. We can point to some most important:

- Adapting the existing algorithms of clustering to the task of clustering of groups.
- Finding the heuristic rules that allow to decrease the number of possible solutions.
- Developing the methods for creating the new properties and relationships.

The first one is a key challenge. Solving this problem would make it possible to use our method in the tasks of the automatic ontology building and the knowledge acquisition. With the support of evolutionary algorithm, we would synthesize an optimized agent system possessing the hierarchical structure of knowledge. That knowledge may be relatively easily verified and adapted into other systems.

References

1. Michalski, R.S., Carbonell, J.G., Mitchell, T.M. (eds.): Machine Learning: An Artificial Intelligence Approach, vol. 2. Morgan Kaufmann, San Mateo (1986)
2. Michalski, R.S.: Inferential Theory of Learning and Inductive Databases, In: UQAM Summer Institute in Cognitive Sciences, Montreal (June 30-July 11, 2003)
3. Muggleton, S.H., De Raedt, L.: Inductive logic programming: Theory and methods. Journal of Logic Programming 19/20 (1994)
4. Muggleton, S.H., Feng, C.: Efficient induction of logic programs. In: Proceedings of the Workshop on Algorithmic Learning Theory, Ohmsha, Tokyo (1990)
5. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach, 3rd edn. Prentice-Hall, Englewood Cliffs (2010)
6. Tadeusiewicz, R., Ogiela, M.R.: Medical Image Understanding Technology. Studies in Fuzziness and Soft Computing, vol. 156. Springer, Heidelberg (2004)

A Generic Approach to Design and Querying of Multi-purpose Human Motion Database

Wiktor Filipowicz[1], Piotr Habela[1],
Krzysztof Kaczmarski[2], and Marek Kulbacki[1]

[1] Polish-Japanese Institute of Information Technology, Warsaw, Poland
{wiktor,habela,kulbacki}@pjwstk.edu.pl
[2] Warsaw University of Technology, Warsaw, Poland
k.kaczmarski@mini.pw.edu.pl

Abstract. The advancement of motion recording, analysis and synthesis techniques, together with the standardization of respective data formats, constitutes a solid foundation for research based on motion data sets in the context of various disciplines. However, most of the motion data sets available offer groups of files as acquired from the motion capture systems. The problem with such data is that it usually represents a single viewpoint, its context varies and is more or less implicit, and the lack of functionality implemented atop of such data set limits the data analysis and search potential. This encourages us to look at this problem domain from the database management systems (DBMS) state of the art point of view. In this paper, we outline some important aspects of applying a DBMS to motion data with the aim to provide a highly universal, extensible, shareable and searchable resource. To avoid being locked into a specific area of application, we take a very abstract view of the data and attempt to assure, a versatility and genericity of the resulting system.

Keywords: Motion capture, multimedia databases, structured data, extendibility, database management.

1 Challenges for a Motion Data Analysis System

Human motion databases (HMDB) have become very popular, gathering large amounts of often synchronized motion capture streams and other motion related data [1]. Captured motion data sets have multiple attributes and dimensions and different lengths for even similar motions. There is no direct correspondence between data of two different motions. The challenges in HMDB include feature selection relevant for motion comparison, indexing methods for fast retrieval of preferred motions, cleaning and compressing data sets, identification of base characteristics of human motion especially of outliers. To classify and recognize motion data correctly we often represent each motion stream as feature vector. Query formulation by feature extraction gives only statistically accurate results based on the frame rate or the position of the motion stream but rare on exact object matching. Nevertheless initial preprocessing and generalization by

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 105–113, 2010.

reduced representations can dramatically accelerate retrieval times. Forbes and Fiume [2] relying on dynamic time warping (DTW) demonstrate how clustering and dimensionality reduction can improve retrieval time. To the same effect Chiu et al. [3] use self organizing map (SOM) for clustering and separately index segments of the body to improve retrieval times. For retrieval of motions from examples in iterative way Kovar and Gleicher [4] preprocessed locally optimal time alignments of motions for whole HMDB. Muller et al. [5] employ binary features to represent various poses and demonstrate how to produce radically faster motion retrieval. Quoted retrieval examples represent a group of retrieval problems where complete example motion is provided as a query. Query based methods in multimedia databases are not able to represent user's subjective intentions perfectly. Second group of retrieval problems in HMDB require from user browsing through database for expected motion. Liu Ren [6] explores representation for easier data visualization in browsing-style interface. Assa et al. [7] present a visual synopsis of motions for elegant database browsing. Sakamoto and colleagues [8] demonstrate visual interface for motion selection by identifying key frames from posture map obtained from SOM algorithm. Third group of retrieval problems relies on retrieval from small set of controls to capture user's intention more reliably. Liu et al. [9] propose hierarchical database comprised of local linear models for fast retrieval of a character pose from a sparse marker configuration.

Motion stream as relatively new type of multimedia require considering temporal and spatial multiatribute motion nature. This variability of data processing requirements demands for a a large degree of flexibility in underlying data management.

At the highest level we recognize four main groups of requirements for the motion database system:

- to offer efficient, reliable, flexible and robust data storage;
- to allow querying, analysis and data retrieval;
- to organize cooperation of different groups of users;
- to enable the extensibility features.

Each of them influences the others and therefore should always be discussed together. An ideal system allows users to store any data related to human motion domain: database records, files of any type and size. The data and files are searchable in different ways, by built-in tools and also by custom procedures and filters. The system guarantees data protection. Hence, to enable cooperation, it should able to anonymize data to assure privacy of captured subjects. An ideal system gathers groups of users and allows them to share not only data but also applications working on the stored data. Community-developed tools for motion analysis can be shared and executed by many users in different situations. This kind of system is similar to community web portals where all content can be shared and is in fact a product of a common development effort.

2 Existing Examples of Public Motion Data Sets

We investigated publicly available data sets including the most important and well known [1], [10], [11], [12]. In all cases the phrase "database" means that there is data that can be accessed (read) by users. Only one of them facilitates uploading of data using a web interface. Sharing and community interactions are therefore usually impossible. All those datasets offered data stored as single or grouped files containing video, motion and also other synchronized data.

Large motion data collections can also constitute a challenge in the medical field. Simple tools for motion data acquisition and processing may fall short when mining in the existing archive is considered. Motion capture lab workstations are clearly not capable of storing massive amounts of records for an extended period of time. A hospital based motion capture lab which we analyzed is forced to frequently archive its motion data using DVD media. Hundreds of disks are stored and an index of their content in terms of patient data location is maintained. Although patient records become documented, the research potential of such a big data set becomes void due to poor search and analysis capability. A dedicated data server with huge capacity could definitely help, but only if files are stored under a professional database system control instead of raw files.

Animeeple [13] is a character animation tool which can import 3D files and publish characters and animations for online sharing or sale. A number of popular formats is supported by its import and export features. The available characters and animations can be combined from different sources. Content can be shared for free or sold by registered sellers and profiles are used to share information about individual content providers. As a 3D modeling and animation tool, Animeeple aims to be an easy to use application targeted mostly at less advanced users and anyone interested in publishing their content for others to use.

Gaitabase [14] is a database of medical records and data captured by gait analysis systems with a web interface. It includes a publicly shared sample data set which can be used as part of a demo. The main purpose of the system is to store gait data files, provide query mechanisms and display the results for statistical analysis. Gaitabase is a relational database with tables storing spatial-temporal files. The search capabilities offer a way to define filters based on search criteria with multiple conditions and logical operators. Processed filters return records from the table of patients as search results. The analysis functionality built into the system makes it possible to plot gait data, select variables and compare sets of data.

As can be seen, there are significant limitations of storing motion data set files “as is”, without a dedicated data management tool. Then we could observe how diverse the expected features of such a tool can be depending on the area of application. For multimedia authoring, the identification of motion kind, interoperability in terms of different formats provided and capability to edit motion are essential. For the medical applications the primary subject of queries are the attributes describing a patient and data processing is a more read-only fashion. In both cases efficient ways of querying and easy access to retrieved data stored inside motion data files are desirable.

3 Required Features of the Database

Useful features, high efficiency and stability, place the DBMSs in general right in the center of our attention. Confronting our requirements with the features traditionally associated with multimedia data management [15], we need to consider the following:

– Handling large, raw data files,
– Need for automatic metadata extraction for imported items,
– Some aspects of temporal querying,
– Need for custom query optimizations (including specific indexing),
– Specific presentation tools and user interface design.

On the other hand, contrary to some other multimedia applications, no real-time challenge occurs. Moreover, the spatial aspect of data – although it obviously applies, will probably not be handled explicitly by the database mechanism itself. Moreover, a specific challenge is the need for high degree of universality and openness for new data formats, metadata elements, search criteria, transformation and analysis routines, connected with the ability of secured shared access and resource contribution.

Data storage and querying
Despite the amount of information is high, the system's core data elements are not expected to form big graphs and would not be the subject of intensive navigation (i.e. much of the complexity would remain opaque to the database schema structures). The foundational set of entities is considered to form a following hierarchy: **Performer** – identifies a subject of motion; either a human or an individual created using 3D modeling software; **Session** – identifies motion capture sessions for a single performer; **Trial** – identifies trials which may be done by a performer during a single session; **Segment** – describes a fragment of the trial sequence distinguished based on some criteria of interest. Different kinds of segmentation of a single segment are possible, especially facing the multiple areas of applications. At each level of this hierarchy it should be possible to assign instances to categories and to equip them with custom attributes. Good support for bulk data is needed to assure fast and reliable motion data storage and retrieval as most of the data will remain stored inside such large objects. One of the most important factors for data querying is the intuitiveness of the data structure as perceived by the database user. From this point of view the following elements need to be covered with a suitable abstraction:

– Links between categories, performers, sessions, trials and files to ease and simplify the formulation of queries referring to more than one of those entities;
– Named generic attributes and their values should be as easy for browsing and querying as regular, statically defined features;
– Features that can be determined by processing motion data files (here called "analytical features") should be available just like plain, stored attributes;

- Services providing derived data that describe certain entity should also be available in a similar way as regular attributes.

The above assumptions result in a concept of object-like, mostly hierarchical structure, having a significant amount of behavior is built in and where the distinction between attributes and operations may be hidden from the user.

Extensibility features and user cooperation
When analyzing extensibility we must focus on the two main aspects of the system: data storage and data querying. In case of the first one we expect that the data stored in the system does not necessarily have to be limited to predefined attributes or file formats. Therefore, no limitation on file type is acceptable. Similarly, a fixed set of attributes for entities is just a starting point from which particular users may start building their own information representation. Summing up, the contributable features would include custom attributes, named queries (filters) and behavior extending data analysis and import / export capabilities. Those behavioral extensions require the following:

1. Standard interfaces to existing data and calculable features, which will be used by new components;
2. Connection points for new components and their safe execution control;
3. A descriptor file specifying the semantics and configuration.

Filters, being a more lightweight construct, only require designing appropriate syntax assuring easy retrieval of existing filters for invocation as well as for modification. For extended cooperation, the system should allow to distribute common activities, including the creation of behavioral parts (e.g. queries or analytical features). This requires many parts traditionally perceived as second class citizens of the system to be promoted to first class citizens. This calls for a new paradigm for user interfaces, query processing and security. These issues are broader discussed in the next section. Security is perceived as an important requirement in the system, especially where medical records are involved. The privileges may need to be differentiated both vertically (against the types of data) and horizontally (particular subsets of data). Hence the access rights pattern goes beyond simple data ownership criteria and would need to be controlled by the application. At the same time, for flexibility and performance, the database content cannot be encapsulated too aggressively against the behavioral extensions.

4 Extensible Architecture for Database and Its Services

Reconciling the genericity and approachability requirements makes it necessary to establish three levels of abstraction:

- **implementation**, including the implementation of internal structures for generic features and resource representations as well as access optimization;
- **design**, outlined by programming and configuration interfaces for administrator and extension developers;

- **use**, where built-in and extension features, both stored and calculated are provided in a uniform way for querying and retrieval.

Here we focus on the second level, perceiving it pivotal for the overall design.

Conceptual view of the generic data

The features considered led us to the following conceptual design, expressed in fig. 1 as a UML class diagram. Note that its complexity would be hidden from regular user, who would actually deal with quite straightforward hierarchy of the resource data.

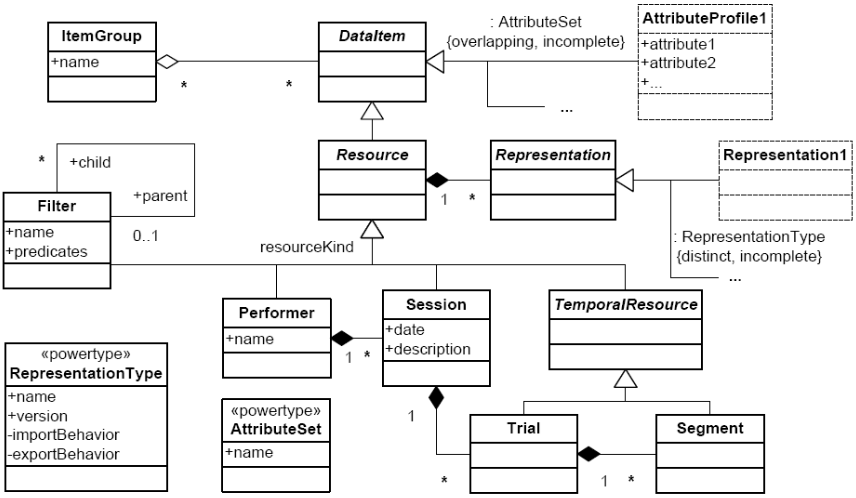


Fig. 1. Motion database structure as perceived by its user

The foundation is formed by the core conceptual entities to be presented to the user: the Performer... Segment hierarchy. We call them Resources. Two of them: Trial and Segment possess temporal characteristics, that are to be supported by queries and when extracting temporal fragments of those resources. The synchronization between different representations would be the substantial value of the system (allowing to investigate the relationships between particular channels). Apart from the resources, we need to explicitly deal with their Representations. Although ideally the representations could be perhaps transparent to this conceptual structure, in fact we need to include them also as the subject of extensible metadata mechanism to support various formats, levels of precision etc. The extension mechanism makes the structure open to new data types in two main roles:

- Item metadata attribute sets - establishing virtual subclasses for resource, whose visibility can be differentiated on a per user basis. For simpler classification a similar, more intuitive notion of Item Group is also available.

- Resource representations - allow supporting new types of stored files or calculated, derived documents.

Behavior for extendible features
The data content extensions of the database entail the need for dynamically incorporating respective behavior:

- New areas of application may bring not only new stored attributes, but also new calculable features. Though made available to the user in a uniform way with the stored attributes, they need a way to connect their calculation algorithms to the query evaluation and reporting routines.
- New resource representations would usually require specifying import behavior that would be capable of automatically extracting their content-dependent metadata and writing it explicitly into the database.
- If a given resource representation is not materialized, it will need submitting an algorithm of producing it based on other resource data (export behavior).
- New topics of research can produce recurring query predicates, so naming and storing them could be useful for sharing or reuse. This requires a relatively simple expression language that could be represented by database-stored Filter entities.

The presence of computation-intense virtual features and derived resource representation makes it also necessary to introduce the postponed result collection query and data retrieval API operation variants.

Plug-ins and sandboxes
Plug-in based extensibility is a commonly accepted approach for enriching software functionality. However, this is mostly used for desktop or server applications under strict administrator control. A user granted with administrator privileges is responsible for any problems caused by a faulty plug-in. Also intranet users pose less risk compared to global network users. If a community open server is to allow users to execute their own code, it must assure a proper level of security:

1. No data can be seen by plug-ins which are not granted with certain privileges;
2. No data or server application can be destroyed by a plug-in code;
3. A plug-in cannot directly or indirectly influence execution of other plug-ins;
4. No plug-in can consume server resources beyond a given level;
5. No plug-in can cutoff normal server communication lines.

Therefore, we consider guarding each executed plug-in in a sandbox, which will be controlled by dedicated server processes. A sandbox-based control means that a process inside is fairly free to perform different tasks, create other processes, data files, etc. if only it obeys given rules (for example, disk occupation or processor load). If a process is trying to break from the shelter it is immediately terminated. Removing its sandbox includes deleting of any intermediate data, rolling-back transactions, etc. The server must offer appropriate interfaces for defining new plug-ins, configuring them, sharing code among other users, controlling their execution, retrieving results. A system similar to [16] could be used for these tasks.

5 Conclusions

We presented here a vision of a generic motion capture database system offering extended features like data sharing and user cooperation, server-side query processing and analysis of motion data, user-defined plug-ins, generic custom defined attributes and openness for a community of different types of users. There has been no similar system created so far and we believe that implementation and deployment of this solution will influence human motion capture society. To realize this vision, several aspects of transparency need to be achieved to provide users and developers with:

- generic metadata attributes which are as simple in querying as the statically defined ones,
- calculable metadata and resources accessible like stored data,
- access rights control factored out of the plug-ins code,
- resources management protecting the performance and availability.

The system seems to be complicated at the server side and since it is a scientific solution, at the beginning it will be implemented in an inventive and incremental way. Small portions will be published for a closed group of users who will test it, collect remarks and help to define the next steps. The process of polishing all requirements and enabling high level of security and reliability will take years, but already the partial results might be useful for large groups of users.

Acknowledgements

This paper has been supported by the project ,,System with a library of modules for advanced analysis and an interactive synthesis of human motion” co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme - Priority Axis 1. Research and development of modern technologies, measure 1.3.1 Development projects.

References

1. Sigal, L., Balan, A., Black, M.: Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. IJCV 87(1), 4–27 (2010)
2. Forbes, K., Fiume, E.: An efficient search algorithm for motion data using weighted pca. In: 2005 ACM SIGGRAPH/Eurographics SCA, pp. 67–76 (July 2005)
3. Chiu, C.Y., Chao, S.P., Wu, M.Y., Yang, S.N.: Efficient content-based retrieval for motion capture data. Journal of Visual Communication and Image Representation 15, 446–466 (2004)
4. Kovar, L., Gleicher, M.: Automated extraction and parameterization of motions in large data sets. ACM Transactions on Graphics 23(3), 559–568 (2004)
5. Demuth, B., Müller, M., Eberhardt, B.: An information retrieval system for motion capture data. In: Lalmas, M., MacFarlane, A., Rüger, S.M., Tombros, A., Tsikrika, T., Yavlinsky, A. (eds.) ECIR 2006. LNCS, vol. 3936, pp. 373–384. Springer, Heidelberg (2006)

6. Ren, L.: Statistical Analysis of Natural Human Motion for Animation. PhD thesis, CMU (2007)
7. Assa, J., Caspi, Y., Cohen-Or, D.: Action synopsis: pose selection and illustration. ACM Transactions on Graphics 24(3), 667–676 (2005)
8. Sakamoto, Y., Kuriyama, S., Kaneko, T.: Motion map: image-based retrieval and segmentation of motion data. In: 2004 ACM SIGGRAPH/Eurographics SCA, pp. 259–266 (July 2004)
9. Liu, G., Zhang, J., Wang, W., McMillan, L.: A system for analyzing and indexing human-motion databases. In: Proceedings of the ACM SIGMOD, pp. 924–926 (2005)
10. Gaeorgia Tech: HID Database, http://www.cc.gatech.edu/cpl/projects/hid
11. CMU: Carnegie-Mellon Mocap Database, http://mocap.cs.cmu.edu
12. CASIA Gait Database, http://www.sinobiometrics.com
13. Animeeple Animation Software, http://www.animeeple.com
14. GAITABASE, http://gaitabase.rch.org.au
15. Prabhakaran, B.: Multimedia Database Management Systems. Springer, Heidelberg (1996)
16. Global Alliance WS-Resource Framework, http://www.globus.org/wsrf

Surveillance Video Stream Analysis Using Adaptive Background Model and Object Recognition

Paweł Forczmański and Marcin Seweryn

West Pomeranian University of Technology
Faculty of Computer Science and Information Technology
Żołnierska Str. 49, 71–210 Szczecin, Poland
{pforczmanski,mseweryn}@wi.zut.edu.pl
http://en.wikipedia.org/wiki/West_Pomeranian_University_of_Technology

Abstract. The paper presents an idea of real-time video stream analysis which leads to the detection and tracking of suspicious objects that have been left unattended, which is one of the most crucial aspects to be taken into consideration during the development of visual surveillance system. The mathematical principles related to background model creation and object classification are included. We incorporated several improvements to the background subtraction method for shadow removal, lighting change adaptation and integration of fragmented foreground regions. The type of the static regions is determined by using a method that exploits context information about foreground masks, significantly outperforming previous edge-based techniques. Developed algorithm has been implemented as a working model involving freely available OpenCV library and tested on benchmark data taken from real visual surveillance systems.

1 Introduction

Typical visual surveillance system is aimed at gathering information about certain phenomena in order to execute or suggest proper actions, especially when it comes to situations dangerous to human health, life and property or environment. In practice, a visual surveillance is implemented using a closed-circuit television system (CCTV) and is used to maintain close observation of a person or a group of people [1–3]. It is typical to visual surveillance systems that static camera is aimed at one fixed point in space which does not move during observation. The focal length of camera lens is constant too. Background modeling is one of the most important tasks of modern intelligent surveillance systems. Its efficiency determines the capabilities of the whole system, hence it often incorporates very sophisticated methods of image analysis. Since the scenes observed by the CCTV cameras are not static, the process of background separation has to take into consideration many different environmental conditions, such as variable lighting [4], atmospheric phenomena and changes caused by humans or vehicles. There are many methods of background modeling based on different principles, but all

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 114–121, 2010.

of them can be divided into two main categories: pixel-based and block-based approaches. The former class of methods analyses each individual pixel in the image, while the latter divides an image into segments (often overlapping). For each pixel or segment several features are calculated used in the classification (into background and foreground objects). Each detected object is then tracked. Most of the tracking methods assume that the movement is constant and the direction does not change in a considerable way [5, 6], however it is questionable in a case of tracking human beings. The last stage of an intelligent surveillance system involves object recognition. The adequate recognition method depends mainly on the system purpose, hence access control requires some biometric identification, traffic control involves ”reading” license plates, and outdoor security systems implement human detection and tracking. In a system presented in [7] each detected object is classified by mean area it occupies, while the authors of [8] detect humans using average human body proportions and size. These approaches are simple to implement but their application is limited.

The paper is organized as follows. First, system outline is presented, followed by background model, object detector, tracker, and then object classifier. Finally, some experimental results and summary are given.

2 Proposed System

2.1 Overview

We present an algorithm that analyses video sequence in order to find potentially security relevant events connected with abandoned objects in public facilities. There have been many such algorithms proposed over last years, however the problem seems to be still unsolved. The main problem is proper object classification and tracking. Proposed system automatically detects foreground objects, classifies them and tracks their trajectories. We distinguish between three classes of objects that can be present at the scene: moving humans, possibly dangerous objects and other objects, e.g. vehicles and other moving objects. Let us assume that a video sequence produced by a system consists of a set of frames covering rather static background and moving objects in front of it. These objects (e.g. humans or vehicles) move in a stepwise manner, which is caused by reduced frame rate (often equal or less than 25 frames-per-second). In cases when camera or observation point move strongly or there are rapid changes of illumination, the sequence should be cut into segments which are analyzed independently.

Proposed system consists of adaptive *Background Model*,*Object Tracker* and *Classifier*. Background elimination is the main task of the background model. Foreground regions are segmented from the present frame in respect to the previously learned background model. Sequence of detected foreground blobs are afterwards sent to the Tracker, which updates tracked objects and tries to predict their next positions. At the beginning of tracking, every newly created object is classified by the Object Classifier and its label is stored in Object Tracker. In the end bounding boxes of tracked objects are marked on the output frame. A simplified scheme of the proposed system is depicted in Fig. 1.

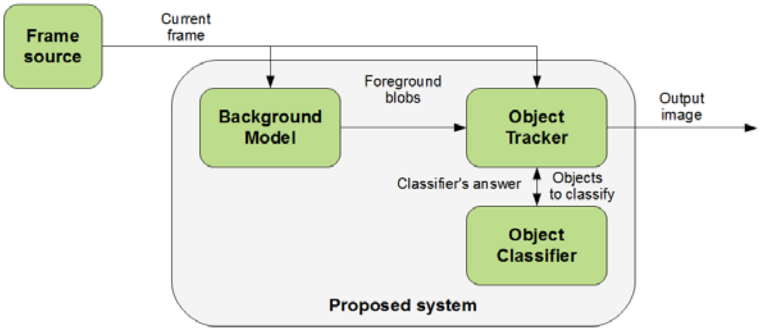


Fig. 1. Overview of the proposed system

2.2 Background Model

We employ adaptive background model similar to the one proposed in [4], in which every pixel is modeled by a mixture of Gaussians in R, G and B channels. In our approach we use five Gaussians distributions for each pixel in the frame, which increases the robustness of the model in the comparison to the one presented in [9].

Every pixel value from the current frame is checked against the existing Gaussians in the corresponding position in the background model. If no match is found the least probable Gaussian is replaced by the new distribution with the current value as a mean. Afterwards, weights of existing Gaussians are updated according to the following: weights of distributions that do not correspond with the new pixel value are decreased, while the weights of distributions that suite the new pixel value are increased. Parameters of unmatched distributions remain the same. The parameters of the distribution which matches the new observation are updated as follows:

$$\mu_t = (1-\rho)\mu_{t-1} + \rho X_t, \tag{1}$$

$$\sigma_t^2 = (1-\rho)\sigma_{t-1}^2 + \rho(X_t-\mu_t)^T(X_t-\mu_t), \tag{2}$$

$$\rho = \alpha\eta(X_t|\mu_k,\sigma_k), \tag{3}$$

where X_t is a new pixel value, η is a Gaussian probability density function, α is a learning rate, μ and σ are distribution parameters, and $\rho \in \langle 0,1\rangle$.

Next, each weight of each distribution is updated according to the following:

$$\omega_t = \begin{cases} (1-\alpha)\omega_{t-1}+\alpha & \text{if a pixel fits the distribution} \\ (1-\alpha)\omega_{t-1} & \text{otherwise} \end{cases} \tag{4}$$

Background subtraction operation results in an image mask of possible foreground pixels which are grouped using connected components. Its main drawback is the lack of shadow suppressing functions. Shadow cast by the moving objects can be a root of serious problems. It can greatly affect localization and detection of objects present at the scene. In proposed system we use a shadow detection method based on [10]. In our approach the algorithm detects shadows in Hue Saturation Value (HSV) color space which is proven to correspond closer to human perception of color than the RGB color space. Used method bases on the observation that casted shadow darkens the point while chrominance of shaded and open regions does not vary much. This assumption can be easily depicted in HSV color space and turns detecting shadow problem into checking three simple conditions:

$$SP(x,y) = \begin{cases} 1 \text{ if } \alpha \leq \frac{I^V(x,y)}{B^V(x,y)} \leq \beta \wedge (I^S(x,y) - B^S(x,y)) \leq \tau_S \\ \quad \wedge |I^H(x,y) - B^H(x,y)| \leq \tau_h \\ 0 \text{ otherwise} \end{cases} \tag{5}$$

where SP is a new binary mask of a foreground blob, α, β, τ_h, τ_s are empirically chosen parameters (according to heuristic approach described in [10]), I is a current frame, B is a background image and H, S and V superscripts point which component of HSV pixel should be used, x and y are pixel coordinates in the image. Initial condition checks the first part of the assumption that shadow darkens the region, whereas next two conditions check if hue and saturation for shaded and open regions do not vary too much. The results of shadow suppressing function are presented in Fig. 2 and Fig. 3.

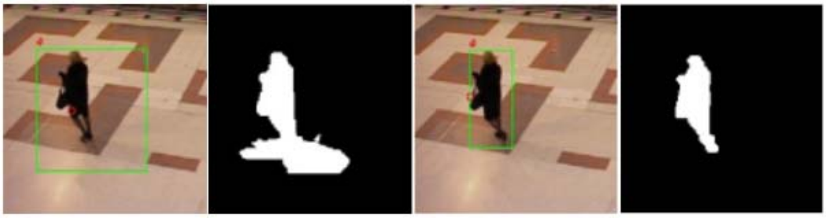


Fig. 2. Detected moving object and shadow elimination results

2.3 Object Tracker

Detected foreground objects are tracked by Object Tracker. Each tracked object is associated with an information about its bounding box, last and predicted position, serial number and group label. Object Tracker uses Kalman filters to predict further positions of objects [11]. Objects are tracked from frame to frame in a stepwise manner using predicted position or correspondence between their last position and foreground blobs. We use a Kalman filter-based approach instead of very popular Mean-Shift algorithm [7] since it is independent on the object appearance, hence it makes it possible to track objects that are partially occluded or changing its silhouette over time.

Matching detected foreground blobs to tracked objects is done by constructing association matrix similar to the one proposed in [12]. For all pairs of blobs and objects we measure Euclidean distance from predicted position of object to the center of the foreground blob, if the foreground blob includes predicted position. If foreground region does not include predicted position, but it intersects with last bounding box of the tracked object we measure distance from the center of bounding box to the center of the blob. If the foreground blob does not include predicted position and does not intersect with the object, distance between them is equal to plus infinity and will be ignored. After evaluating distance between all pairs of blobs and objects, objects are updated using blobs which are closest to them. If a blob has no matched object, a new tracked object is created. On the other hand when the object has not matched any foreground blob for several frames, it is removed.

2.4 Object Classifier

Most of the approaches presented in the literature use single classifier in order to determine the type of moving objects. In our approach we propose to use two independent classifiers that use different features. The first one employs Histograms of Orientated Gradients which was proposed in [13], while the second classifier is based on Haar-like features and was described by Viola and Jones in [14]. We propose to use a cascade of classifiers, in which every object sent to classifier module is classified first by the HOG-based classifier and then, in case of negative result, is classified once more by the Haar-based classifier. This method of classification aims to suppress number of false-negatives and reduces the number of computations.

Images of objects sent to the HOG classifier are divided into small rectangular regions, called cells. For each cell, histogram of gradient orientations is evaluated. Afterwards, cells are grouped into overlapping larger spatial groups, called blocks. To improve performance, every block is contrast normalized. The Final descriptor of the object is a vector of all normalized cell responses from all of the blocks. In the end, classification of descriptor is done by the soft margin linear Support Vector Machine (SVM).

Haar-like features are rectangular regions divided into smaller rectangles which are labeled white or black. A value of a single feature is equal to a difference between sum of pixels which lie within white and black rectangles. Single Haar-like feature is a weak classifier and cannot be used to recognize complex objects like human silhouette. To improve its performance and create the strong classifier, the Adaboost algorithm is used to create cascades of simple Haar-like features. In the cascade, subsequent feature is evaluated only if answer of the previous feature equaled the expected value, otherwise classification is finished with the negative label. Finally, the system marks and labels detected and classified objects. The color of the bounding box is set accordingly to the object's label: *green* - for objects classified as people, *red* - for potentially suspicious objects. All other objects, which have not been unequivocally classified are marked with *blue* rectangles. Sample markings can be seen in Fig. 3.

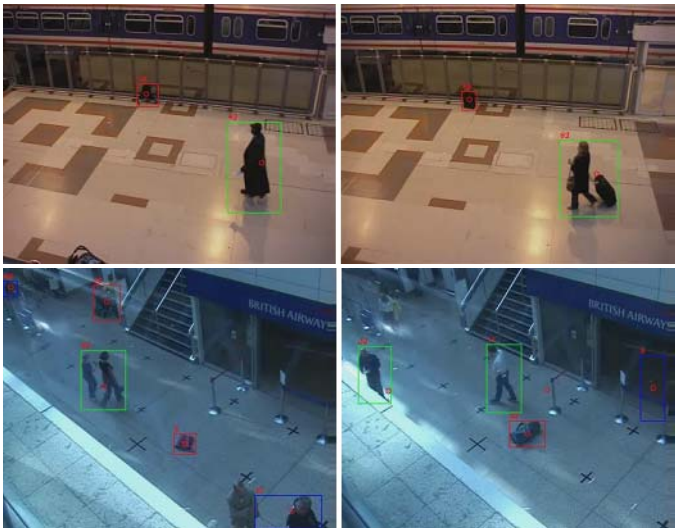


Fig. 3. Example detection presented on PETS 2006 (upper row) and PETS 2007 (lower row) data. Left luggage is marked with red rectangle, while moving humans are marked green; false objects are marked with blue rectangles.

2.5 Experimental Results

The system described in the previous section has been implemented in the C++ using OpenCV library [15]. At a resolution of 360 × 288 pixels, We reached 5 frames per second on an average 1,74GHz PC which is an acceptable result for rather low-end hardware. The implemented system was evaluated on the PETS 2006 and PETS 2007 benchmark data, recorded at busy railway station in London [1] and Glasgow Airport [2]. The aim of the PETS scenarios is to detect left luggage items in the public space. Video sequences contain people moving alone or in groups across the platform in case of PETS 2006 and people loitering at the airport in case of PETS 2007. Every scenario contains also the luggage drop-off event. The parameters for testing the algorithm was set according to the Table 1.

We define a success as a correctly detected and classified left luggage item, as a half-success we define a correctly detected yet misclassified luggage. We define fail as an undetected left luggage item. During the tests we used sequences from camera number 3, which supplies the optimal view of the scene (see sample results in Fig. 3). Detailed result of tests are shown in Table 2.

During the tests we have observed the spurious detections in the tinted glass wall, caused by moving nearby objects. Furthermore, our Object Classifier in several cases misclassified objects which appeared behind the tinted glass. Both misclassification problems were caused by small size and reduced contrast of those objects and can be solved with proper calibration of camera's exposure parameters.

Table 1. Parameters of the algorithm

Stage	Parameter name	Value
Background model	Learning iterations	200
	Number of Gaussians	5
	Minimal weight	0.70
	Maximal deviation	6
	Minimal object size	245
	Initial weights	0.05
	Initial distribution variance	30
	Distance between objects	58
	Scaling factor	0.5
Shadow elimination	Minimal shadow intensity ratio α	0.4
	Maximal shadow intensity ratio β	0.99
	Maximal saturation distance τ_S	10
	Maximal hue distance τ_H	50

Table 2. Results of tests

Dataset	Scenario	Result	Comment
PETS 2006	S1-T1-C	Success	-
	S2-T3-C	Success	-
	S3-T7-A	Success	Luggage item is not abandoned.
	S4-T5-A	Success	Luggage blob is merged with blob of person standing next to it.
	S5-T1-G	Half-success	Lugguage item is detected but misclassified
	S6-T3-H	Success	-
	S7-T6-B	Success	-
PETS 2007	S00	Success	-
	S01,S02,S03	Not tested	Not enough data for model
	S04	Success	-
	S05	Success	Luggage item is not abandoned.
	S06	Half-success	Luggage blob is merged with blob of persons standing next to it.
	S07	Success	-
	S08	Half-success	Luggage item is misclassified

3 Summary

In the paper we presented a model of an intelligent visual surveillance system. It incorporates adaptive background model, object tracker and sophisticated object classifier utilizing Haar-like features and SVM. Developed method has been tested on real-world data and chosen approach proved to be robust. In opposition to the other similar projects it is insensitive to quick light variations and shadows cast by moving people. Its field of application is very wide. The most obvious are security surveillance systems at airports and railway stations for detecting left-luggage, but it can be also utilized in the shopping centers and trade fairs for counting customers in the specific areas.

References

1. Thirde, D., Li, L., Ferryman, J.: Overview of the pets2006 challenge. In: Ninth IEEE International Workshop on Performance Evaulation of Tracking and Surveillance, PETS 2006 (2006)
2. Ferryman, J.M. (ed.): Proceedings Tenth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2007), Rio de Janeiro, Brazil (October 2007)
3. Regazzoni, C.S., Fabri, G., Vernazza, G.: Advanced Video-Based Surveillance Systems. Springer, Heidelberg (1999)
4. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 1999), vol. 2 (1999)
5. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(5), 564–577 (2003)
6. Welch, G., Bishop, G.: An introduction to the kalman flter, course 8. In: SIGGRAPH (2001)
7. Li, L., Ma, R., Huang, W., Leman, K.: Evaluation of an ivs system for abandoned object detection on pets 2006 datasets. In: Ninth IEEE International Workshop on Performance Evaulation of Tracking and Surveillance, PETS 2006 (2006)
8. del Rincon, J.M., Elias Herrero-Jaraba, J., Gomez, J.R., Orrite-Urunuela, C.: Automatic left luggage detecion and tracking using multi camera ukf. In: Ninth IEEE International Workshop on Performance Evaulation of Tracking and Surveillance, PETS 2006 (2006)
9. Tian, Y., Feris, R.S., Hampapur, A.: Real-time detection of abandoned and removed objects in complex environments. In: IEEE International Workshop on Visual Surveillance (in Conjunction with ECCV 2008), Marseille, France (2008)
10. Cucchiara, R., Grana, C., Piccardi, M., Prati, A., Sirotti, S.: Improving shadow suppression in moving object detection with hsvcolor information. IEEE Intelligent Transportation Systems, 334–339 (2001)
11. Piovoso, M., Laplante, P.A.: Kalman filter recipes for real-time image processing. Real-Time Imaging 9(6), 433–439 (2003)
12. Fengjun, L., Xuefeng, S., Bo, W., Singh Vivek, K., Ramakant, N.: Left-luggage detection using bayesian inference. In: Ninth IEEE International Workshop on Performance Evaulation of Tracking and Surveillance (PETS 2006), pp. 83–90 (2006)
13. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, pp. 886–893. INRIA, IEEE Computer Society, Washington (2005)
14. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 511–518 (2001)
15. Open Computer vision library (2009), http://sourceforge.net/projects/opencvlibrary/

Nonlinear Multiscale Analysis of Motion Trajectories

Bartosz Jabłoński and Marek Kulbacki

Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008 Warszawa, Poland
{bjablonski,kulbacki}@pjwstk.edu.pl

Abstract. Multiscale analysis initially proposed in image processing domain, turned out to be an effective tool in many other areas. It has been also proposed for motion data trajectories parametrized in quaternion space. In this paper we introduce novel approach, which is the framework for non-linear scale space analysis of quaternion trajectories. In the paper theoretical basis has been proposed and numerical validation has been presented.

1 Introduction

Effective methods for automatic data analysis and processing are becoming important part of many existing systems. Real world data consists of some information, which is stored and available in different scales. Multiscale analysis has been shown to be effective in image processing and other domains. This paper addresses the issues of using multiscale of signal processing methods in analysis of motion capture. We analyzed the possibility of preparation the generalized multiscale approach for quaternion trajectories basing on non-linear and anisotropic scale space generation. This approach leads to better control over characteristic of motion sequence.

The main goal of the presented work is to provide a flexible framework for describing motion data in different scales. This kind of approach is very important for the analysis of meaningful features of the motion. Specifically this approach is part of the more complex system for human motion analysis. The proposed solution is the basis to build multiscale human motion model. The possibility to extract meaningful features of motion trajectory is crucial to detect human motion regularities and anomalies.

The rest of the paper is organized in the following order. In the second chapter we present papers related to multiscale analysis and processing. Chapter three introduces shortly the manner of motion trajectories quaternion representation. The actual concept of nonlinear multiscale approach is presented in chapter four. Chapter five contains numerical results of multiscale analysis. The paper is summarized with the notes on further works in chapter six.

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 122–130, 2010.

2 Related Work

Scale space has become an important tool and description model for computer vision and image processing. One of its first applications involved building a pyramid of images, which are sampled at different scale and simplified using Gaussian kernel [2] . The concept of signal expansion by convolution with Gaussian masks over a continuum of sizes has been introduced in [12] with more formalized approach proposed in [8]. Authors of the paper [10] propose nonlinear image scale space filtering, which was later extended into fully anisotropic model proposed in [11]. Recent works focus on the problem of automatic scale selection and meaningful features selection [5].

Multiscale analysis has also been used in area of motion processing, for example in a task of motion warping [13]. Later it has been formalized in the framework of multiresolution motion analysis in [7] and practical applications proposed in [4].

The presented paper introduces non-linear framework for scale space generation of motion trajectories. It extends the approach from [7] proposing the diffusion based model, which can lead to better features preservation in higher order scales.

3 Motion Trajectories Representation

We use quaternions for parametrization of orientation space. Quaternions form a group under multiplication called the Hamilton algebra defined as $\mathbb{H} = \{t + xi + yj + zk | t, x, y, z \epsilon R\}$ with the three imaginary units i, j, k. One of the most important advantages of this parameterization in computer graphics is that it does not result in singularities found at other approaches [9]. For any unit quaternion q, the operation qpq^{-1} performs a rotation of the vector p about the direction q. The composition and inversion of rotations correspond to multiplication and inversion in $\mathbb{H}$, yielding easy and direct formulas. The natural distance measure for quaternions is the length of a geodesic line on a four-dimensional unit sphere $\mathbb{S}^3$, which is given by:

$$d(q_1, q_2) = 2 \min \left(\| \log \left(q_1^{-1}, q_2 \right) \|, \| \log \left(q_2^{-1}, q_1 \right) \| \right) \tag{1}$$

Exponentiation q^t is defined as

$$q^t = \exp\left(t \log q \right) \tag{2}$$

and is commonly used within spherical-linear interpolation

$$q_p = q_1 \left(q_1^{-1} q_2 \right)^p \tag{3}$$

It leads to a smooth interpolation between two consecutive orientations by means of a time-dependent rotation of constant angular velocity. In this paper we consider quaternion trajectory as a discrete representation of motion in the space of rotation. Each trajectory element q_i corresponds to the orientation in space for the discrete time moment i.

4 Nonlinear Multiscale Analysis

Multiscale motion analysis proposed in [7] has interesting properties similar to corresponding concept of image pyramids defined in [2]. In the following paragraph we will propose new approach, which is based on non-linear equivalent of diffusion process for quaternion trajectory introduced in [3].

Original multiscale approach is based on the convolution of Gaussian kernel. The same result is obtained by calculating the solution of partial differential equation (PDE) of the diffusion process defined as:

$$\begin{cases} \frac{\partial u(x,t)}{\partial t} = \Delta u\left(x,t\right) \\ u\left(x,0\right) = u_0\left(x\right) \end{cases}, \tag{4}$$

where x is n dimensional vector of variables, t is the time variable for the diffusion process and u_0 is the initial condition. The solution of this equation is the convolution of the initial condition with the Gaussian kernel K of standard deviation $\sqrt{2t}$:

$$u(x,t) = \begin{cases} u_0(x) & \text{for } t = 0 \\ \left(K_{\sqrt{(2t)}} * u_0\right)(x) & \text{for } t > 0 \end{cases}, \tag{5}$$

Image is often considered as two dimensional function of brightness and is taken as initial condition for equation (5). The solution of the equation at time t is the result of scale space generation, where t is the parameter of generated continuous space.

4.1 Nonlinear Diffusion Process

Nonlinear and Anisotropic Extension. In this paper we will propose the model for multiscale quaternion trajectories representation. It's based on PDE diffusion approach as the starting point for constructing generalization of quaternion trajectory nonlinear scale space.

Perona and Malik proposed in [10] the application of anisotropic (actually non-linear) process based on the equivalent of equation (4):

$$\frac{\partial u(x,t)}{\partial t} = div\left(c(x,u,t)\nabla u\right) = c(x,u,t)\nabla^2 u + \nabla c \nabla u. \tag{6}$$

The nonlinearity in this case is introduced by the parameter c, which controls the speed of the diffusion process. Supposing, that it is non-linear, symmetric function of local edge presence estimator, the following sample form is proposed:

$$c(x,u,t) = g\left(\|\nabla u(x,t)\|\right) = \frac{1}{1 + \left(\frac{\|\nabla u(x,t)\|}{K}\right)^2}, \tag{7}$$

where K is the contrast parameter described in [10]. This model is non-linear and the parameter c describes the influence of local features. Fully anisotropic model has been later introduced by Weickert in [11].

Generalization for Quaternion Trajectories. The above formulation relates to the continues process, which is computed in the discrete form using finite difference scheme or equivalent numerical approach. Due to the nature of quaternion space, similar solution could not be directly applied to the specified problem.

Making use of the discrete quaternion differentials suggested in [1] and quaternion spherical interpolation properties we derive the following discrete procedure:

$$q_i^{(k+1)} = q_i^{(k)} \left((q_i^{(k)})^{-1} \left(q_{i-1}^{(k)} \left((q_{i-1}^{(k)})^{-1} q_{i+1}^{(k)} \right)^{1/2} \right) \right)^{\hat{c}} , \tag{8}$$

where the upper index relates to the discrete time variable of the solution (iteration) and the lower index indicates the element number in the trajectory sequence. Parameter $\hat{c}$ controls the speed of the diffusion process. One might observe the similarity of the scheme (8) to the model introduced in [6]. Our solution is the equivalent of numerical scheme for PDE integration, whereas the cited approach results from the energy minimization approach.

The proposed nonlinear trajectory approach is based on Perona-Malik equation (6). For two dimensional data the equation can be solved by the following numerical approximation:

$$u^{k+1} = u^k + \frac{\lambda}{4} \left[c_N \nabla_N u + c_S \nabla_S u + c_W \nabla_W u + c_E \nabla_E u \right]^k , \tag{9}$$

which is four-point (four directions:N, S, W, E) implicit scheme.

Basing on the generalization approach used for linear scheme, we obtain the following approximation of the solution for quaternion trajectory:

$$\tilde{q}_i = q_i \left(q_i^{-1} q_{i-1} \right)^{\tilde{c}_P} \exp \left[0.5 \log \left(\left(q_i \left(q_i^{-1} q_{i-1} \right)^{\tilde{c}_P} \right)^{-1} \left(q_i \left(q_i^{-1} q_{i+1} \right)^{\tilde{c}_N} \right) \right) \right] , \tag{10}$$

considering the following definition for diffusion coefficients:

$$\tilde{c}_P = g \left(\| E_P \| \right) = \lambda g \left(\left\| 2 \log \left(q_i^{-1} q_{i-1} \right) \right\| \right) , \tag{11}$$

$$\tilde{c}_N = g \left(\| E_N \| \right) = \lambda g \left(\left\| 2 \log \left(q_i^{-1} q_{i+1} \right) \right\| \right) . \tag{12}$$

The physical meaning of the two coefficients corresponds to four directional diffusion coefficients of Perona-Malik model. Using similar stopping function g the process speed is decreased in areas, where the meaningful changes of signal are higher. In this case changes are considered in the sense of difference between consecutive quaternions in the trajectory - namely c_N and c_P.

Finally additional dumping coefficient c_A is introduced, which helps to control the speed of the process along with the following transformation:

$$q_i^{(k+1)} = q_i \left(q_i^{-1} \tilde{q}_i \right)^{c_A} . \tag{13}$$

It is possible to consider higher order features similarly as it is used for anisotropic image process. In the next step we introduce anisotropic factor in the sense of considering variable coefficient depending on higher order features like tangential angular acceleration (responsible for the change of local curvature of the trajectory). This kind of directional (as a contrary to scalar) dependency is described by the function of dumping coefficient:

$$\tilde{c}_A(q,t) = \lambda_A g(\kappa(t)), \tag{14}$$

where κ_i is the measure of the change of local curvature describing the difference on unit vectors for local angular velocities:

$$\kappa_i = \left\| \frac{\log\left(q_i^{-1} q_{i+1}\right)}{\|\log\left(q_i^{-1} q_{i+1}\right)\|} - \frac{\log\left(q_{i-1}^{-1} q_i\right)}{\|\log\left(q_{i-1}^{-1} q_i\right)\|} \right\|^2 \tag{15}$$

The coefficient tends to zero for increasing value of the local curvature measure. That means that the diffusion in this area of the trajectory is decreased. As a result the meaningful local feature of signal change is preserved. Non-linear and anisotropic transformations are referenced as F_N and F_A respectively.

4.2 Scale Space Data Model

In the paper [7] the method of multi-resolution trajectory representation has been proposed corresponding to the classical image pyramid multiscale [2]. It down samples the trajectory along with simplification. We propose to represent signals in the form of multiscale construction, where it is simplified, but not down sampled and hence representing infinite scale space [10]. We also introduce nonlinear simplification scheme for the reduction operator, which provides important scale space properties.

We denote the original trajectory signal in the fine scale as $q^{(0)}$. The signal in the i-th scale (into the coarse direction) is denoted as $q^{(i)}$.

Operation of anisotropic trajectory reduction is described by the statement:

$$q^{(i+1)} = F_A(q^{(i)}). \tag{16}$$

Similarly the displacement signal is defined as

$$d^{(i+1)} = q^{(i+1)} \ominus q^{(i)} = (q^{(i+1)})^{-1} q^{(i)}, \tag{17}$$

where $\ominus$ is the general symbol for the difference between two trajectory elements interpreted in our approach as a quaternion difference.

The generation of the signal in scale of order k (meaning, that k discrete consecutive elements of the scale space have been defined) we define as a transformation composition of the following type:

$$q^{(k)} = F_A(q^{(k-1)}) = F_A(...F_A(F_A(q^{(0)}))...), \tag{18}$$

where for each scale k corresponding displacement map $d^{(k)}$ is computed. Two series of scales for $i = 1..k$ build the scale space of order k of the original signal.

Similarly we reconstruct the original signal $q^{(0)}$ in the recursive procedure similar to [7]:

$$q^{(0)} = q^{(1)} \oplus d^{(1)} = (q^{(2)} \oplus d^{(2)}) \oplus d^{(1)} = (((q^{(k)} \oplus d^{(k)}) \oplus d^{(k-1)}) \cdots) \oplus d^{(1)}, \quad (19)$$

but without expansion operation. The reconstruction operator $\oplus$ is defined as

$$q^{(i)} \oplus d^{(i)} = q^{(i)} d^{(i)}. \quad (20)$$

In our method the usage of the expansion operator is not necessary. That means, that the reconstruction operator (20) can be extended with additional local scale parameter p_i:

$$q^{(i)} \oplus d^{(i)} = q^{(i)} \exp(p_i \log d^{(i)}). \quad (21)$$

Hence extended reconstruction procedure will also take vector P as an input parameter defined as

$$P = [p_0, p_1, ..., p_i, ..., p_k], \quad (22)$$

where each element p_i is the interpolation coefficient for extended reconstruction parameter in the range $[0, 1]$. For $p_i = 1$ one obtains the reconstruction of the original scale of signal for order i. Assigning different values to the parameter allows us to control the reconstruction process and to enhance or limit the influence of the scale of given order.

5 Numerical Results for Scale Space Analysis

We have performed number of experiments, but only two results can be presented here due to the limited space. We base our experiments on the trajectory representing real data of human motion (called *Run to Walk*) recorded in one of the motion capture sessions. For the simplicity we are considering the trajectory of only one bone from the hierarchy, namely the right knee. For comparison we have generated linear and non-linear scale spaces - selected scales of the trajectory are presented on figure 1.

As one can observe, the testing trajectory contains couple of meaningful sections - points where shape of trajectory changes significantly. The rate of change for the motion speed and trajectory curvature are the most important features used for classification and distinguishing different motions.

For both linear and nonlinear scale spaces the signal is simplified for coarse scales. However for anisotropic version the simplification process is stopped in areas where meaningful features of the trajectory exist (e.g. change of the curvature). Linear transformation tends to oversimplification of the space scale signal (see third column on figure 1). It is characteristic feature of linear approach. Similarly two dimensional image scale space tends to the average value of brightness over original matrix. These properties are clearly visible on figures 3 and 2. Charts present scalar value of displacement component for each scale.

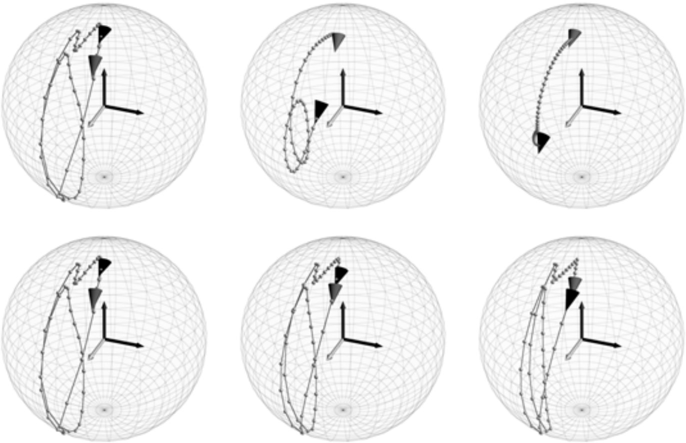


Fig. 1. Scale space generation for the test signal: first row - linear transformation, second row - anisotropic transformation; columns correspond to scale orders 0, 16, 64

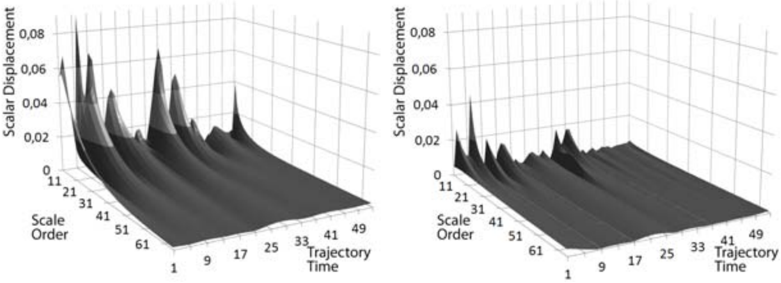


Fig. 2. Scalar component of displacement signal for scale space up to order 64. Left chart - linear transformation, right chart - anisotropic transformation

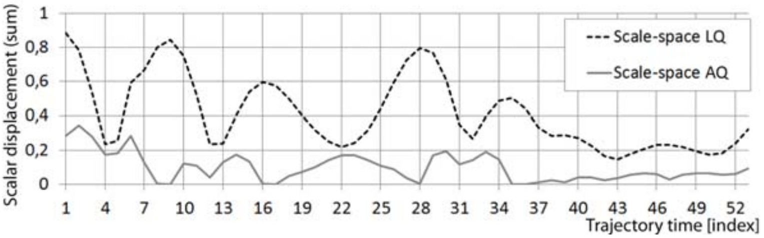


Fig. 3. Comparison of the sum of scalar displacements for scale space of maximum order 64 - linear and anisotropic scale spaces

These figures show that there are sections of the trajectory, where the diffusion is stopped with anisotropic transformation. These sections can be easily detected by analyzing the sum of scalar value for the displacement signal - see for example elements number 7, 16 or 28 (figure 3). The lower the value, the more transformation was limited in this area. Using this approach it should be possible to detect of meaningful features, by analysis of the signal in scale-space.

Due to the limited space, we are not able to provide detailed analysis of the proposed approach. However presented results clearly show the difference between regular linear scale space and newly proposed non-linear approaches.

6 Conclusions and Future Works

In the paper we have proposed a new method of non-linear and anisotropic scale space analysis for quaternion trajectories. We have introduced anisotropic transformation, which lets to simplify the signal and to enhance its characteristic features. Multiscale data model has been proposed to analyze the quaternion trajectory at different levels of detail.

Theoretical foundations have been introduced and the results of basic numerical experiments discussed. We have used real motion capture data as the source signal for numerical experiments. We also conclude that the proposed solution can be the starting point to create effective algorithms for the detection of important meaningful features of the trajectory.

Acknowledgements

This paper has been supported by the project “System with a library of modules for advanced analysis and an interactive synthesis of human motion” co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme - Priority Axis 1. Research and development of modern technologies, measure 1.3.1 Development projects.

References

1. Barr, A.H., Currin, B., Gabriel, S., Hughes, J.F.: Smooth Interpolation of Orientations with Angular Velocity Constraints using Quaternions. Computer Graphics 26(2) (1992)
2. Burt, P.J., Adelson, E.H.: The Laplacian Pyramid as a Compact Image Code. IEEE Trans. on Communications 31(4), 532–542 (1983)
3. Jablonski, B.: Anisotropic filtering of multidimensional rotational trajectories as a generalization of 2D diffusion process. Multidimensional Systems and Signal Processing 19(3-4), 379–399 (2008)
4. Kim, T., Lee, J., Shin, S.Y.: Robust Motion Watermarking based on Multiresolution Analysis. Computer Graphics Forum 19(3) (2000)
5. Laptev, C., Lindeberg, T.: Interest point detection and scale selection in space-time. In: Proc. Scale-Space Methods in Computer Vision, pp. 372–387. Springer, Heidelberg (1994)

6. Lee, J., Shin, S.Y.: Motion Fairing. Computer Animation,136–143 (1996)
7. Lee, J., Shin, S.Y.: Multiresolution Motion Analysis with Applications. In: International Workshop on Human Modeling and Animation, Seoul, pp. 131–143 (2000)
8. Lindeberg, T.: Scale-Space Theory in Computer Vision. Kluwer Academic Publishers, Netherlands (1994)
9. McCarthy, J.M.: Introduction to Theoretical Kinematics. MIT Press, Cambridge (1990)
10. Perona, P., Malik, J.: Scale-Space and Edge Detection Using Anisotropic Diffusion. IEEE Trans. On Pattern Analysis and Machine Intelligence 12(7) (1990)
11. Weickert, J.: Anisotropic Diffusion in Image Processing. B.G. Teubner, Stuttgart (1998)
12. Witkin, A.: Scale-Space Filtering. In: IJCAI, pp. 1019–1022 (1983)
13. Witkin, A., Popovic, Z.: Motion Warping. In: Proc. SIGGRAPH 1995, Los Angeles (1995)

Matlab Based Interactive Simulation Program for 2D Multisegment Mechanical Systems

Henryk Josiński[1,2], Adam Świtoński[1,2], Karol Jędrasiak[1],
Andrzej Polański[1,2], and Konrad Wojciechowski[1,2]

[1] Polish-Japanese Institute of Information Technology
Aleja Legionów 2 41-902 Bytom, Poland
{hjosinski,aswitonski,kjedrasiak,apolanski,kwojciechowski}@pjwstk.edu.pl
[2] Silesian University of Technology
Akademicka 16 44-100 Gliwice, Poland
{Henryk.Josinski,Adam.Switonski,
Andrzej.Polanski,Konrad.Wojciechowski}@polsl.pl

Abstract. This paper presents principles of designing multisegment mechanical system. Represented is a model of a single segment, its extension in form of a couple of segments and the final construction – a fragmentary and simplified silhouette of a human form named the biped. This paper describes procedure of construction of the biped's digital model using Matlab package. It also discusses test run of a single experiment and algorithm of the calculations realized in the single step of integration.

1 Introduction

Scientists create physical and mathematical models of movement of a whole human form or its specified parts. Mathematical models enable creation of digital models to conduct computer simulation. Model testing makes it possible to analyse decomposition of forces and torques in a non-invasive way.

The purpose of this research was to build a simple multisegment movement model in order to determine the principles of adding the successive segments and how to avoid their break-up. This is how the silhouette named biped was created. Its digital model created using Matlab package should enable to carry out simulation of the biped's movement.

Literature related to this subject is extensive. Analysis of contents of a movement model was included in [9]. The dynamics of planar human body motion, solved with a non-iterative matrix formulation was presented in [1]. Worth mentioning is also item [3] of the bibliography where Brubaker et al proposed a model based on the *Anthropomorphic Walker* [5, 6], a physics-based planar model with two straight legs, a single torsional spring and an impulsive collision model. The Anthropomorphic Walker is simple, as it only exhibits human-like gaits on level ground. Brubaker et al [2] introduced also the *Kneed Walker* – a complex model of bipedal locomotion based on biomechanical characterizations of human walking [7]. It has a torso and two legs with knees and ankles. It is capable of exhibiting a wide range of plausible gait styles. A mathematical model of the swing phase of walking was presented in [8].

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 131–138, 2010.

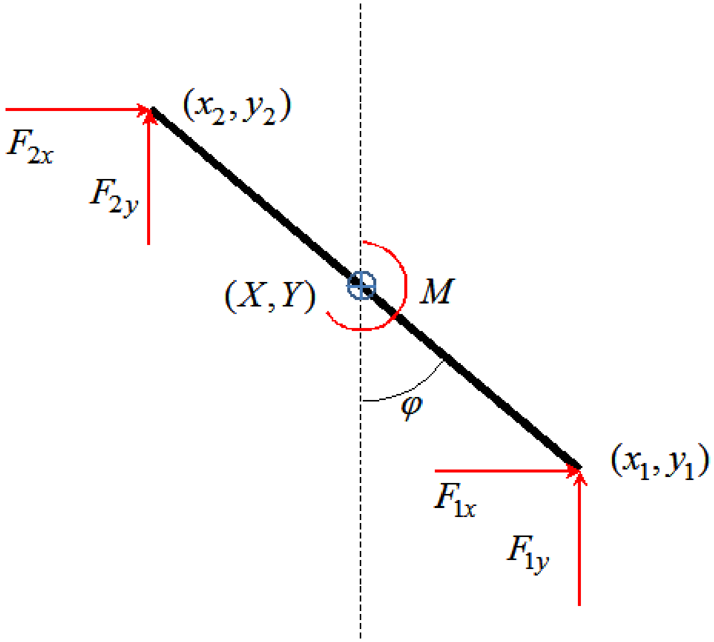


Fig. 1. Decomposition of forces for a single segment

2 Single Segment Movement Model

Let's consider single segment with the mass m and length l, where centre of the mass is positioned in the point (X, Y), its ends are at the points (x_1, y_1), (x_2, y_2) and the angle of vertical inclination is φ (Fig. 1). Let's presume action of external forces: F_1 with components F_{1x} and F_{1y}, F_2 with components F_{2x} and F_{2y} and external control torque M. The following equations describe the segment's state:

$$\ddot{X} = \frac{F_{1x} + F_{2x}}{m} \tag{1}$$

$$\ddot{Y} = \frac{F_{1y} + F_{2y}}{m} - g \tag{2}$$

$$\ddot{\varphi} = \frac{l\left[(F_{1x} - F_{2x})\cos\varphi + (F_{1y} - F_{2y})\sin\varphi\right]}{2I} + \frac{M}{I} \tag{3}$$

where g represents acceleration of gravity and I is segment's moment of inertia relative to its axis of rotation crossing centre of the mass $I = ml^2/12$.

Equations allow to determine acceleration's vertical and horizontal component and also angular acceleration. State variables are represented by the centre of the mass coordinates (X, Y), vertical and horizontal components of centre of mass velocity $\left(\dot{X}, \dot{Y}\right)$, angle φ and angular velocity $\dot{\varphi}$.

Up to date segment's ends' components are calculated using simple trigonometric equations:

$$x_1 = X + \frac{l\sin\varphi}{2} \qquad y_1 = Y - \frac{l\cos\varphi}{2} \tag{4}$$

$$x_2 = X - \frac{l\sin\varphi}{2} \qquad y_2 = Y + \frac{l\cos\varphi}{2} \tag{5}$$

Differentiation allows to calculate vertical and horizontal components of segment's ends' velocity:

$$\dot{x}_1 = \dot{X} + \dot{\varphi}\frac{l\cos\varphi}{2} \qquad \dot{y}_1 = \dot{Y} + \dot{\varphi}\frac{l\sin\varphi}{2} \tag{6}$$

$$\dot{x}_2 = \dot{X} - \dot{\varphi}\frac{l\cos\varphi}{2} \qquad \dot{y}_2 = \dot{Y} - \dot{\varphi}\frac{l\sin\varphi}{2} \tag{7}$$

Successive differentiation leads to formulae describing segment's ends' acceleration:

$$\ddot{x}_1 = \ddot{X} + \ddot{\varphi}\frac{l\cos\varphi}{2} - \dot{\varphi}^2\frac{l\sin\varphi}{2} \qquad \ddot{y}_1 = \ddot{Y} + \ddot{\varphi}\frac{l\sin\varphi}{2} + \dot{\varphi}^2\frac{l\cos\varphi}{2} \tag{8}$$

$$\ddot{x}_2 = \ddot{X} - \ddot{\varphi}\frac{l\cos\varphi}{2} + \dot{\varphi}^2\frac{l\sin\varphi}{2} \qquad \ddot{y}_2 = \ddot{Y} - \ddot{\varphi}\frac{l\sin\varphi}{2} - \dot{\varphi}^2\frac{l\cos\varphi}{2} \tag{9}$$

Equations describing state of the segment allow to formulate general relationship between accelerations of the segment's ends and external influences – forces and a control torque M:

$$\begin{bmatrix} \ddot{x}_1 \\ \ddot{y}_1 \\ \ddot{x}_2 \\ \ddot{y}_2 \end{bmatrix} = M_{coef} \begin{bmatrix} F_{1x} \\ F_{1y} \\ F_{2x} \\ F_{2y} \\ M \\ 1 \end{bmatrix} \tag{10}$$

Symbol M_{coef} represents coefficients' matrix determined by the formulas (8), (9) allowing for equations (1), (2), (3).

3 Model of Couple of Segments Movement

The case of 2 joined segments requires doubling of the state variables' set applied for the case of a single segment. Decomposition of forces and torques is shown on Fig. 2 (index U denotes the upper segment whereas L – the lower one).

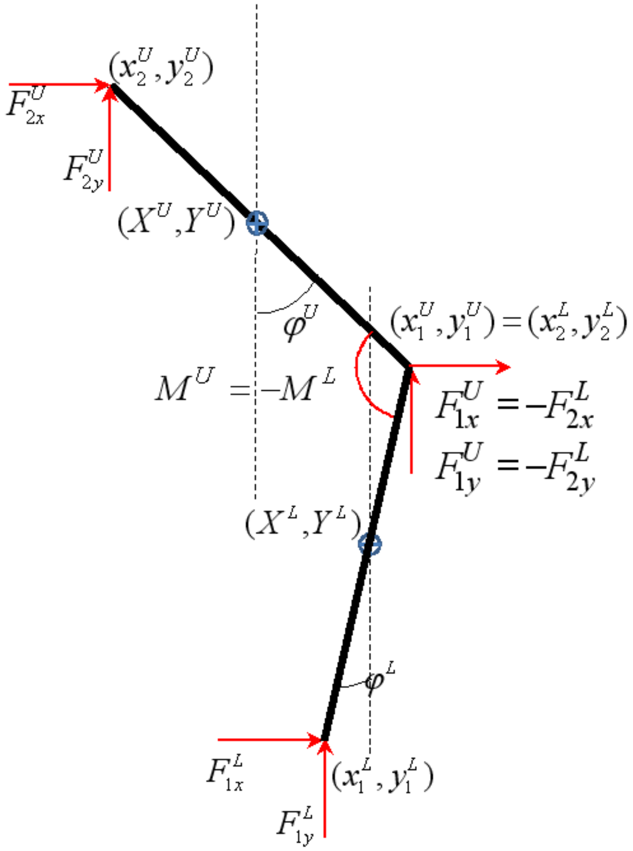


Fig. 2. Decomposition of forces for a couple of segments

For the point of osculation of both segments (a joint) the following dependencies are fulfilled:

$$F_{1x}^U = -F_{2x}^L = F_x^{UL} \tag{11}$$

$$F_{1y}^U = -F_{2y}^L = F_y^{UL} \tag{12}$$

$$M^U = -M^L = M^{UL} \tag{13}$$

where symbols F_x^{UL}, F_y^{UL} were introduced as notation for reaction forces.

In the setup of two segments it is very important to prevent any break-up of the segments. For joint ends of both segments $\left(x_1^U, y_1^U\right)$, $\left(x_2^L, y_2^L\right)$ the following conditions should hold:

$$x_1^U\left(t + \mathit{\Delta} t\right) = x_2^L\left(t + \mathit{\Delta} t\right) \qquad y_1^U\left(t + \mathit{\Delta} t\right) = y_2^L\left(t + \mathit{\Delta} t\right) \tag{14}$$

Expansion of the functions $x_1^U(t)$, $y_1^U(t)$, $x_2^L(t)$, $y_2^L(t)$ into Taylor series leads to following dependencies (with sufficiently small Δt):

$$x_1^U(t+\Delta t) \cong x_1^U(t) + \dot{x}_1^U(t)\cdot\Delta t + \ddot{x}_1^U(t)\cdot\frac{\Delta t^2}{2} \tag{15}$$

$$x_2^L(t+\Delta t) \cong x_2^L(t) + \dot{x}_2^L(t)\cdot\Delta t + \ddot{x}_2^L(t)\cdot\frac{\Delta t^2}{2} \tag{16}$$

$$y_1^U(t+\Delta t) \cong y_1^U(t) + \dot{y}_1^U(t)\cdot\Delta t + \ddot{y}_1^U(t)\cdot\frac{\Delta t^2}{2} \tag{17}$$

$$y_2^L(t+\Delta t) \cong y_2^L(t) + \dot{y}_2^L(t)\cdot\Delta t + \ddot{y}_2^L(t)\cdot\frac{\Delta t^2}{2} \tag{18}$$

Conditions (14) enable to calculate unknown values of reaction forces F_x^{UL}, F_y^{UL}. With this end in view appropriate parts of the coefficient matrices M_{coef}^U, M_{coef}^L should be substituted to the general formula (10) bearing in mind the dependencies (11), (12), (13).

Next stage of the extension of multisegment system is setting of 2 couples of segments into a fragmentary and simplified silhouette of a human form named the biped (Fig. 3).

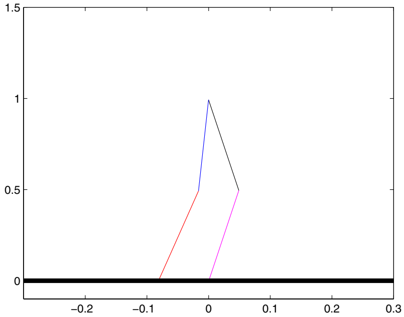


Fig. 3. Biped – example of the multisegment mechanical system

In the biped system there are following reaction forces:

1. From interaction upper leg – lower leg (separately for left and right leg).
2. From interaction upper right leg – upper left leg in the joint *root*.
3. Between the ground and a leg (separately for left and right leg).

The biped's digital model was created by means of the Matlab package and used in simulation experiments.

The goal of the discussed experiment was to simulate a biped's jump down on the ground. The following values of the model parameters were applied: mass of each segment $m = 1$, length of each segment $l = 0.5$, vertical coordinate of the *root* $y_{2L}^U(0) = y_{2R}^U(0) = 2$, angles of vertical inclination of individual segments: $\varphi_L^U(0) = \pi/5$, $\varphi_L^L(0) = 2\pi - \pi/5$, $\varphi_R^U(0) = \pi/6$, $\varphi_R^L(0) = 2\pi - \pi/6$.

4 Test Run of a Single Experiment

Experiment begins with setting of state variables' initial values. Variables' new values are calculated in consecutive moments spaced by the actual value of the integration step. This digital model applied *ode*45 integration method which is one of many Matlab methods for solving ordinary differential equations. It uses 4th and 5th order Runge-Kutta formulas and is based on the Dormand-Prince method [4].

Single integration step determines values of the right sides of the state equations. The calculations are carried out in following stages:

1. Kinetics of individual ends of the segments – calculation of location and velocity.
2. Dynamics of individual ends of the segments – calculation of coefficients' matrices.
3. Forces of reaction with the ground – at the current stage of the research they are set to 0.
4. Reaction forces in the joints joining segments – usage of the Gauss elimination method with the application of coefficients' matrices calculated in the point 2.
5. Dynamics of the centres of the mass of individual segments – calculation of accelerations.
6. Drawing of the actual location of the segments.
7. Placing of calculated values of the right sides of the state equations (velocities as values of the appropriate state variables calculated in the previous integration step and accelerations calculated in the point 5) to the appropriate state variables' derivatives.
8. Integration by means of the method *ode*45 gives new values to the state variables.

Attainment of the limiting value of simulation time ends the experiment. The aim of this investigation was to find a graph depicting distance between joint ends of adjacent segments. Such distance was calculated as Euclidean distance between two points. For a couple of segments from the Fig. 2 the applied formula is as follows:

$$d = \sqrt{\left(x_1^U - x_2^L\right)^2 + \left(y_1^U - y_2^L\right)^2} \tag{19}$$

Graphs acquired by using this formula for left and right knee joints are shown on Fig. 4A and 4B, respectively.

For the joint *root* the formula (19) was modified as follows:

$$d = \sqrt{\left(x_{2L}^U - x_{2R}^U\right)^2 + \left(y_{2L}^U - y_{2R}^U\right)^2} \tag{20}$$

(indexes $2L$ and $2R$ denote upper ends of the upper segments of the left leg and the right one, respectively). Graph acquired for the *root* is shown on Fig 5.

The order of magnitude of the segments' ends' distance read out from the graphs documents the empirical observation that the adjacent segments don't break up.

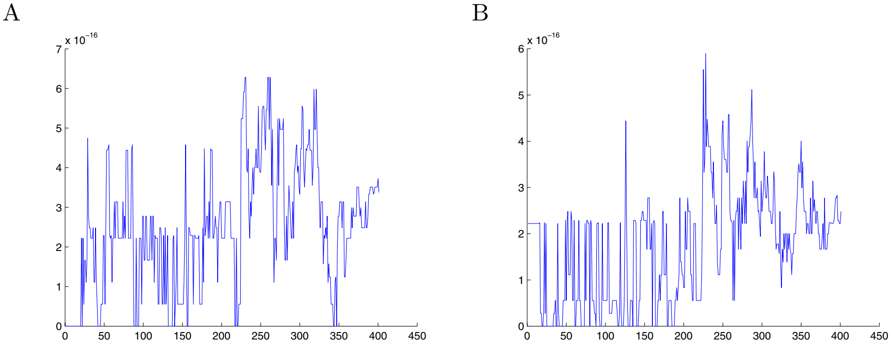


Fig. 4. The distance between common segments ends: A) of the left knee B) of the right knee

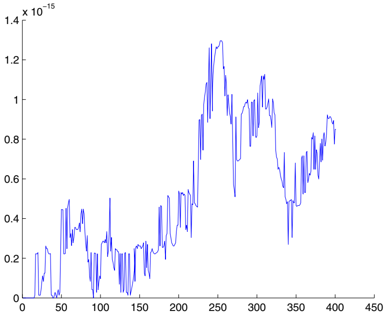


Fig. 5. The distance between common segments ends of the *root*

5 Conclusion

This paper aimed to present the method of construction of a simulation program for 2D multisegment mechanical systems. A model of such a system was created and implemented using Matlab package. Further investigations will include extensions of the model – first of all worth mentioning is the problem of biped's various types of gait. Next plans comprise transfer of the model to 3D space and its numerical stability analysis.

Acknowledgements

This paper has been supported by the project "System with a library of modules for advanced analysis and an interactive synthesis of human motion" co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme – Priority Axis 1. Research and development of modern technologies, measure 1.3.1 Development projects.

References

1. Alciatore, D., Abraham, L., Barr, R.: Matrix Solution of Digitized Planar Human Body Dynamics for Biomechanics Laboratory Instruction. In: Proceedings of the 1992 ASME International Computers in Engineering Conference (1992)
2. Brubaker, M.A., Fleet, D.J.: The Kneed Walker for human pose tracking. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Anchorage (2008)
3. Brubaker, M., Fleet, D.J., Hertzmann, A.: Physics-based person tracking using simplified lower-body dynamics. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis (2007)
4. Dormand, J.R., Prince, P.J.: A family of embedded Runge-Kutta formulae. Journal of Computational and Applied Mathematics 6(1), 19–26 (1980)
5. Kuo, A.D.: Energetics of Actively Powered Locomotion Using the Simple Walking Model. Journal of Biomechanical Engineering (2002)
6. Kuo, A.D.: Dynamic Walking Tutorial. In: NACOB 2008, Ann Arbor (2008)
7. McGeer, T.: Dynamics and Control of Bipedal Locomotion. Journal of Theoretical Biology (1993)
8. Mochon, S., McMahon, T.A.: Ballistic Walking. Journal of Biomechanics 13, 49–57 (1980)
9. Pandy, M.G.: Advanced Computer Modeling of Human Movement. In: Clinical Research Methods in Gait Analysis, Gait CCRE (2006)

Electrooculography Signal Estimation by Using Evolution–Based Technique for Computer Animation Applications

Robert Krupiński and Przemysław Mazurek

West Pomeranian University of Technology, Szczecin
Department of Signal Processing and Multimedia Engineering
26–Kwietnia 10 Str., 71–126 Szczecin, Poland
robert.krupinski@zut.edu.pl, przemyslaw.mazurek@zut.edu.pl
http://rkrupinski.zut.edu.pl
http://media.zut.edu.pl

Abstract. The measurement of electrooculography biosignals (EOG) gives the possibility of eye orientation and eyelid movements (blinking) estimation what is important for realistic eyes motion–capture. In this paper the linear smooth pursuit movements of eyes are considered. The optimization techniques for the estimation of eye signal and the separation of EOG and blinking signals are applied. The reconstruction of both signal is possible by the application of a signal model with a coefficient set, which are the subject of optimization.

1 Introduction

Animation based techniques for virtual avatars (e.g. virtual humans) are based on three methods of motion control. The first method is based on human–assisted animation by using a key–framing technique. This method is very old and is derived from 2D animation by using painting, but creates the possibility of creation of any movements for any character. Unfortunately, this method is both time and cost consuming so other techniques are interesting alternatives. The second method is based on the references shot of a real actor, where the computer vision techniques are used in motion–capture systems.

There are two general variants of motion–capture systems: marker–based and marker–less. In the film industry an optical marker–based motion–capture systems are used most frequently. Such a system uses multiple cameras for the tracking of tens or hundreds of markers (usually an infra–red reflective balls) for the estimation of actor's pose and movements. The third technique is based on the synthetic generation of movements what is used in games and movies and those tools use a key–framing or motion–capture data as the movement components. The adequate eyes movements are very important for the computer generated realistic avatars. A conventional key–framing technique is time consuming for the animation of eyes. The motion capture techniques are not suitable for the tracking of eyes movements. There are some possibilities of eyes movement synthesis [6,10], but with limited correlation to the scene context. There is a good

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 139–146, 2010.

solution for such a case – a dedicated motion–capture system (marker–less) used for the eye movement recordings. These systems are well known in medicine and are named as oculography systems. The obtained estimates could be used mainly for medical purpose [8], but in the last decade a new applications area arose: the ergonomics, the advertisement analysis [17], the human–computer interaction (HCI) systems (e.g. a virtual keyboard [21], the vehicle control [2], the wearable computers [5]), and the video compression driven by eye–interest [11]. Those systems are also applied for the computer animation of eyes, and a good example is the Beowulf movie [8,4,16] where an electrooculography (EOG) system was used successfully.

There are a few types of such systems: the Infra–Red OcculoGraphy (IROG) based on the reflection of IR light, the VideoOcculoGraphy (VOG) based on the video tracking of eye, the ElectroOcculoGraphy (EOG) based on the measurement of biopotentials from a retina–cornea constant voltage source [15,18]. The voltage measurements are an eye angle dependent what is the basic idea of the EOG 1. This source has about $\pm 1[mV]$ [19], which is a very high value in comparison to other biosignals, but it also depends on the numerous factors, for instance, the light conditions [7], the contact between electrodes and skin, which is the source of amplitude instability [12,1].

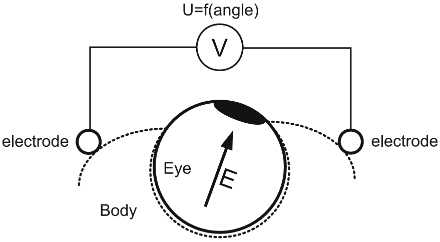


Fig. 1. Model of retina–cornea voltage source and the method of measurement

2 Electrooculography Signals

All oculography methods are interesting for computer animation, where it is very important to combine such systems with motion–capture systems. The IROG requires a special glasses with sensors. The VOG requires a camera, which creates this system difficult to use. One of the most important factor that reveals the advantage of EOG over other oculography methods is the eyes safety and the limitations of actor performance. The IROG or VOG should be located near an eye what is an eye safety problem. A VOG camera located in distance (e.g. 20–50 cm from the face) limits the actor performance. The EOG systems use a few small electrodes (even the flat electrodes could be used) placed on the skin, not necessarily in direct proximity to the eyes what is much safer especially for the action scenes.

There are a few electrode configurations (Fig. 2) and in this paper the 3/4 configuration is assumed: three main electrodes, for two differential measurements: L–U and L–R, and additional fourth reference electrode (REF).

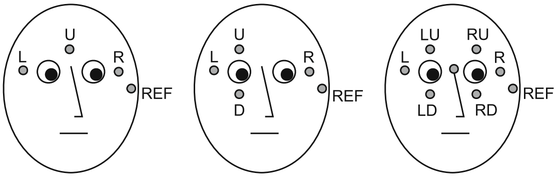


Fig. 2. 3/4, 4/5, and 7/8 electrodes measurement systems

This is a minimal set of electrodes that assumes the symmetry of eyes and blinking performance. A signal from a single differential channel is processed for the estimation of one movement direction of an eye. The combination of two channels gives the possibility of eye orientation estimation.

Two signals are observed by using a single differential channel: an EOG and blinking signals. The blinking signal is usually treated as the distortion of EOG,but it is also an important signal for computer animation. The estimation of time moments of blinking adds realism to the computer generated avatar. The placement of markers on the eyelids is not safe and the biosignal measurement extends the possible performance capture.

The EOG signal is a step like and the time moments with the rapid changes of orientation are named as a 'saccades' [8,9,10,3,13]. The smooth pursuit (SP) movements are linear movements and occurs during an object tracking [8]. There are also another eye movements [8], but they are not considered in this paper. A blinking signal is a pulse like, an additive signal dependent on the electrodes placement. For the 3/4 configuration they are similar to the Gaussian pulses. In this paper it is assumed typical blinking with the short time pulses. There are also the long time period blinking pulses, when an eyelid closes and opens slowly.

3 Estimation of Electrooculography Signals

There are a lot of methods for the separation of EOG and blinking signals and the most typically is a median filtering. This technique is valuable, but it is not sufficient for the separation of blinking pulses in the neighborhood of saccade. In [14] it was proposed another approach based on the optimization algorithm for fitting the synthetic EOG and blinking signals to the real one. Both signals model for the generation of these signals is used, where the coefficients sets consist the parameters for both signals. This approach gives the possibility of signals separation and moreover, the estimation of all signals parameters at one time. The detailed description of the algorithm is in [14]. It is only a selected part explained, which is introduced and in this paper.

The optimization approach uses a set of optimization operations. The substep operations are selected by the random number generators during a single iteration (Fig. 3).

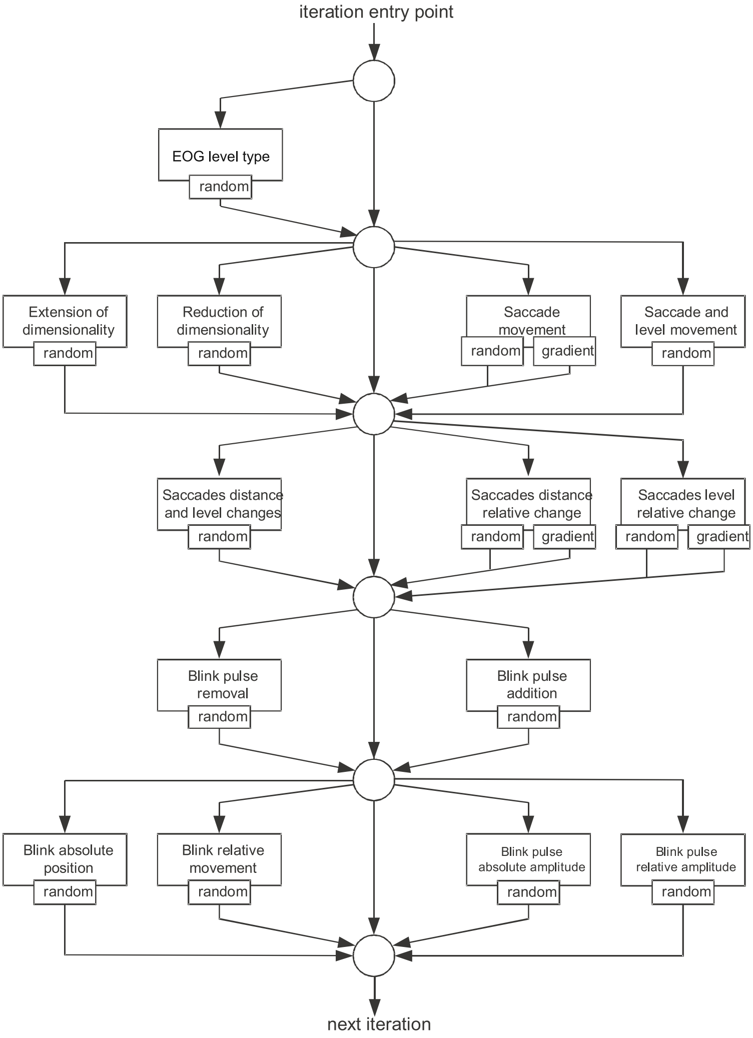


Fig. 3. Processing scheme in the optimization approach

The optimized parameters are divided into two groups: the EOG and blinking signal descriptors. The EOG signal is modeled by the time periods t_i and a constant level value between two saccades s_i. The blinking signal is estimated as a set of blink pulses located at a position b_j and with a corresponding pulse height h_j. The approximated reconstruction of a signal is given by the following formula:

$$\hat{Y} = \sum_i f_S\left(S, i\right) + \sum_j f_B\left(B, j\right) \quad (1)$$

The synthesis of intervals between two saccades is the f_S function and the synthesis of blink pulse is the f_B function. Two indexes i and j correspond to the number of both functions, and they are also a subject to optimization.

For smooth pursuit movement (SP), the extended part of the algorithm, uses a linear interpolation of EOG signal between two neighborhood saccades. An additional flag (l_i) is used in the first part of iteration step (Fig. 3) for the selection of EOG signal approximation type. If this flag is set to zero a step like shape is used and if it is set to one a linear interpolation is used for the smooth pursuit approximation.

The evolutionary (mutation) algorithm is used for the discreet number of parameters, such as the number of blinks and saccades. The mutation and random value are used for most parameters, but a gradient algorithm is also used for some operations such as a saccade level or a blink position finding. There are also more sophisticated operations for the modification of positions and the EOG levels around a saccade for the reduction of convergence problems and the amount of local minima. The algorithm uses the initialization of parameters by a median filter (for the blinking signals) and the segmentation of EOG by using a differential filter for an edge detection [14]. The segmentation algorithm uses the additional filtering of peaks for the limitation of minimal distance between those peaks (the saccades in the original signal). This optimization reduces the computation time a few times.

4 Results for Smooth Pursuit Movement and Optimization Approach

The reduction of MSE is similar for about 300 steps and after those steps the significant reduction of MSE is obtained for the smooth pursuit (about two times lower value). An example signal estimation is show in Fig. 4 for both variants of algorithm.

A very important part of the smooth pursuit supported algorithm is the proper initialization of this part. There is no special initialization such as for the blink positions and levels, the saccades position and the level of signal between the saccades. The most significant part of MSE (Mean Square Error) is related to the EOG shape. The Monte Carlo test could be used for testing the algorithm performance and 1000 test were calculated by using a conventional algorithm without a smooth pursuit (No SP) and the extended version with a smooth pursuit (SP). The mean value for all test is depicted in Fig. 5.

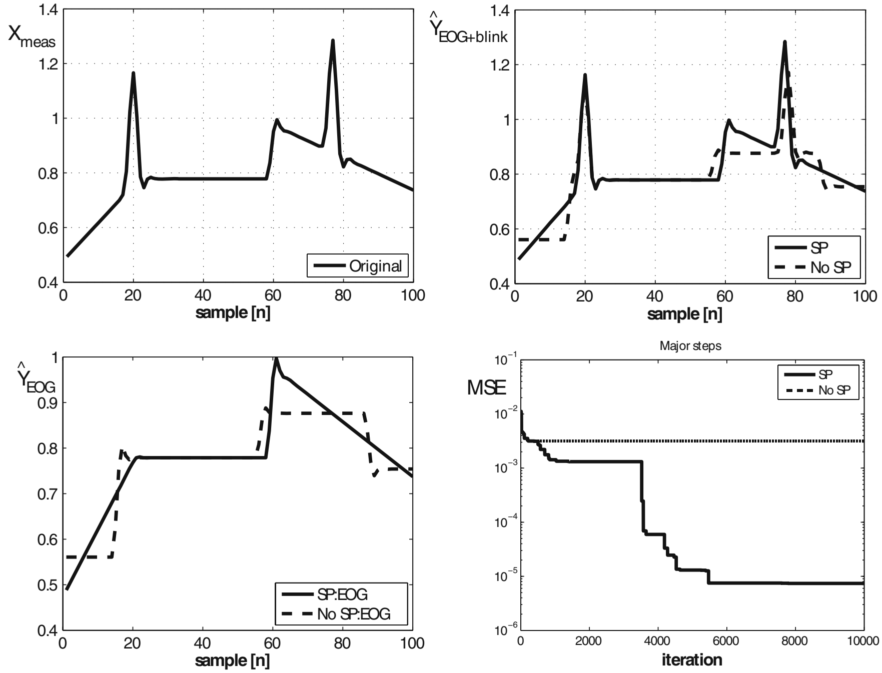


Fig. 4. Example of algorithm performance: original signal (left–top) with smooth pursuit movements, estimated original signal (right–top), estimated EOG signal (left–bottom), and MSE reduction in time (right–bottom)

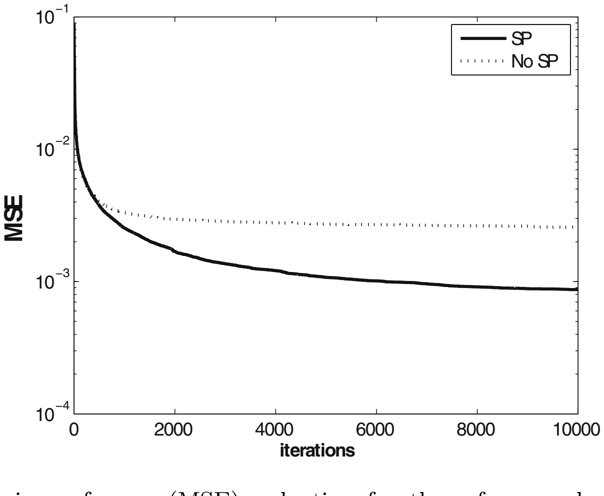


Fig. 5. Comparison of error (MSE) reduction for the reference algorithm without smoooth pursuit (No SP) and algorithm with smooth pursuit (SP)

5 Conclusions

The estimation of eyes orientation and blinking time moments is very important for the application of EOG measurement systems for computer animation. The extended method (including the smooth pursuit movement estimation) leads to the possibility of processing real signals. The proposed algorithm is not a real–time one, but it could be used after actor performance recording. This is sufficient for most applications. A real–time capability with a few second latency could be obtained by using the high performance computing devices (such as VLSI or FPGA chips, GPGPUs, or computer clusters).

Acknowledgments. This work is supported by the UE EFRR ZPORR project Z/2.32/I/1.3.1/267/05 "Szczecin University of Technology – Research and Education Center of Modern Multimedia Technologies" Poland.

References

1. Augustyniak, P.: Przetwarzanie sygnałów elektrodiagnostycznych (textbook in Polish, 244 pages, Uczelniane Wydawnictwa Naukowo–Dydaktyczne AGH, Kraków (2001)
2. Barea, R., Boquete, L., Mazo, M., López, E.: Wheelchair Guidance Strategies Using EOG. Journal of Intelligent and Robotic Systems 34, 279–299 (2002)
3. Becker, W.: The Neurobiology of Saccadic Eye Movements. 1989 Metric. In: Wurtz, R.H., Goldberg, M.E. (eds.), pp. 13–67. Elsevier Science Publishers BV (Biomedical Division) (1989)
4. E.O.G Beowulf DVD 2'nd disc, Warner Brothers (2008)
5. Bulling, A., Roggen, D., Tröster, G.: Wearable EOG goggles: Seamless sensing and context–awareness in everyday environments. Journal of Ambient Intelligence and Smart Environments (JAISE) 1(2), 157–171 (2009)
6. Deng, Z., Lewis, J.P., Neumann, U.: Realistic Eye Motion Synthesis by Texture Synthesis. In: Deng, Z., Neumann, U. (eds.) Data–Driven 3D Facial Animation, pp. 98–112. Springer, Heidelberg (2008)
7. Denney, D., Denney, C.: The eye blink electro–oculogram. British Journal of Ophthalmology 68, 225–228 (1984)
8. Duchowski, A.: Eye Tracking Methodology: Theory and Practice. Springer, Heidelberg (2007)
9. Fleming, B., Dobbs, D.: Animating Facial Features & Expressions. Charles River Media, Hingham (1999)
10. Gu, E., Lee, S.P., Badler, J.B., Badler, N.I.: Eye Movements, Saccades, and Multiparty Conversations. In: Deng, Z., Neumann, U. (eds.) Data–Driven 3D Facial Animation, pp. 79–97. Springer, Heidelberg (2008)
11. Khan, J.I., Komogortsev, O.: Perceptual video compression with combined scene analysis and eye–gaze tracking. In: Duchowski, A.T., Vertegaal, R. (eds.) ETRA 2004 – Proceedings of the Eye Tracking Research and Application Symposium 2004, San Antonio, Texas, USA, March 22-24, p. 57 (2004)
12. Krogh, E.: Normal values in clinical electrooculography. 1. Material, method, methodological investigations and distribution of the potential and time parameters. Acta Ophthalmol (Copenh) 53(4), 563–575 (1975)

13. Krupiński, R., Mazurek, P.: Estimation of Eye Blinking using Biopotentials Measurements for Computer Animation Applications. In: Bolc, L., Kulikowski, J.L., Wociechowski, K. (eds.) ICCVG 2008. LNCS, vol. 5337, pp. 302–310. Springer, Heidelberg (2009)
14. Krupiński, R., Mazurek, P.: Optimization–based Technique for Separation and Detection of Saccadic Movements and Eye–blinking in Electrooculography Biosignals. In: Advances in Experimental Medicine and Biology. Springer, Heidelberg (2010) (in print)
15. Northrop, R.B.: Noninvasive Instrumentation and Measurement in Medical Diagnosis. CRC Press, Boca Raton (2002)
16. Sony Pictures Entertainment, Sony Corporation, Sagar, M., Remington, S.: System and method for tracking facial muscle and eye motion for computer graphics animation, Patent US., International Publication Number WO/2006/039497 A2 (13.04.2006)
17. Poole, A., Ball, L.J.: Eye Tracking in Human–Computer Interaction and Usability Research: Current Status and Future Prospects. In: Ghaoui, C. (ed.) Encyclopedia of Human Computer Interaction, pp. 211–219. Idea Group, USA (2005)
18. Prutchi, D., Norris, M.: Design and Development of Medical Electronic Instrumentation. Wiley, Chichester (2005)
19. Schlgöl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, R., Pfurtscheller, G.: A fully automated correction method of EOG artifacts in EEG recordings. Clinical Neurophysiology 118, 98–104 (2007), doi:10.1016/j.clinph.2006.09.003
20. Thakor, N.V.: Biopotentials and Electrophysiology Measurement. In: Webster, J.G. (ed.) The Measurement, Instrumentation, and Sensors Handbook, vol. 74. CRC Press, Boca Raton (1999)
21. Usakli, A.B., Gurkan, S., Aloise, F., Vecchiato, G., Babiloni, F.: On the Use of Electrooculogram for Efficient Human Computer Interfaces, Computational Intelligence and Neuroscience, Article ID 135629, 5 pages (2010) doi:10.1155/2010/135629

Articulated Body Motion Tracking by Combined Particle Swarm Optimization and Particle Filtering

Tomasz Krzeszowski, Bogdan Kwolek, and Konrad Wojciechowski

Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008 Warszawa
http://www.pjwstk.edu.pl

Abstract. This paper proposes the use of a particle filter with embedded particle swarm optimization as an efficient and effective way of dealing with 3d model-based human body tracking. A particle swarm optimization algorithm is utilized in the particle filter to shift the particles toward more promising configurations of the human model. The algorithm is shown to be able of tracking full articulated body motion efficiently. It outperforms the annealed particle filter, kernel particle filter as well as a tracker based on particle swarm optimization. Experiments on real video sequences as well as a qualitative analysis demonstrate the strength of the approach.

1 Introduction

Human body tracking has many applications, including, but not limited to, visual surveillance, human computer communication and recognizing human activities [1]. One problem of human body tracking is to estimate the joint angles of a human body at any time. This is one of the most challenging problems in the area of computer vision because of self-occlusions, a high dimensional search space and high variability in human appearance. The high dimensionality of the configuration space and the exponentially increasing computational cost are the main challenges in full articulated motion tracking [2]. An articulated human body can be thought of as including at least 11 body parts. This may involve around 26 parameters to describe the full body articulation. By building a mapping from configuration space to observation space, 3d model-based approaches rely on searching the pose space to find the body configuration that best-matches the current observations [3]. Matching such complex and self-occluding model to human silhouette might be especially difficult in cluttered scenes. In monocular image sequence this matching problem is under constrained. The major problems with monocular 3d body tracking arise due to depth ambiguities, movements perpendicular to the image plane and occlusion. Multiple cameras and simplified backgrounds are often employed to ameliorate some of such practical difficulties.

Particle filtering is one of the most popular algorithms for tracking human body motion. After the seminal work [4] the particle filter has been utilized in human motion tracking in [2]. In a particle filter each particle corresponds to

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 147–154, 2010.

some hypothesized set of model parameters. Given the number of parameters needed to describe a realistic articulated model of the human body is larger than twenty, the number of particles of that are required to adequately approximate the underlying probability distribution in the body space might be huge. Hence, a considerable amount of approaches has been proposed to overcome the course of dimensionality inherent in the particle filtering. Given the number of allowable configurations of the human body is restricted by biomechanical constraints, some approaches to human motion tracking are based on learning a low-dimensional representation of the high-dimensional configuration space. Tracking of human motion in such a low-dimensional manifold results in lower numbers of required particles. Gaussian Process Latent Variable Models (GPLVM) [5] provide inverse mapping from the pose space to the latent space. However, manifolds can only be learned for specific activities, such as walking, jumping or running, and it unclear how this approach can be extended to broader classes of human motion.

The likelihood function in human motion tracking can be very peaky with multiple local maxima. In [6], to cope with multiple local maxima the particles are repositioned via a deterministic search in order to attain a better concentration around significant modes of the importance function. A different approach has been proposed in [7] where the promising areas in the probability distribution are identified through iterative mode-seeking using mean-shift. Experiments on real monocular image sequences demonstrated that the algorithm is capable of tracking two arms of upper human body at 7.5 Hz on a laptop computer. Another approach that has gained considerable interest in this type of problems consists in a coarse to fine search on the importance function of the particle filter [2]. Very good results were obtained in a setup with three cameras. As reported in [8] the annealed particle filter has good performance if the images are acquired with frame rate of 60 fps and the tracking performance of such a filter drops when the frame rate is below 30. Another disadvantage of the annealed particle filter is that it often fails to represent multiple hypotheses. In [9] it has been shown that particle swarm optimization outperforms the simulated annealing in terms of accuracy and consistency of the results.

One drawback of using particle filter in human motion tracking is the inability of samples to explore the probability distribution efficiently. This arises because the particles do not move according to their former experience and their relationship to other particles. Hence, they have reduced capability to escape the local minima. Therefore, in this work we propose an algorithm which combines particle swarm optimization (PSO) [10] and particle filtering as an effective way for human motion tracking. The interactions between particles in the course of swarm-based optimization lead to the emergence of global and collective behavior, which allows the particles to gravitate towards the global extremum, whereas the particle filter is responsible for maintaining multi-modal densities.

In the reminder of this paper we briefly outline particle filtering and particle swarm optimization. We then show our algorithm. Afterwards we discuss our results in more detail. Finally, a summary is presented.

2 The Algorithm

Particle swarm optimization is a population based optimization technique, which differs from other evolutionary techniques by inclusion of particle velocity. Particles can be attached to each other by any kind of neighborhood topology represented by a graph. In the fully connected neighborhood topology, which is represented by fully connected graph all particles in the swarm are connected to one another. Each particle in a swarm represents a candidate solution of the problem. With respect to a fitness function, the best location that has been visited thus far by a particle is stored in the particles memory. The fitness values corresponding to such best positions are also stored. Additionally the particles have access to the best location of the whole swarm, i.e. a position that yielded the highest fitness value. A particle therefore employs the best position encountered by itself and the best position of the swarm to move itself toward an optimal value of the objective function.

Each particle i maintains the current position $\mathbf{x}_i$, current velocity $\mathbf{v}_i$, and its best position $\mathbf{y}_i$. For every iteration, the jth-component of particle velocity vector $\mathbf{v}_i$ is updated as follows:

$$v_{i,j} \leftarrow w v_{i,j} + c_1 r_{1,j}(y_{i,j} - x_{i,j}) + c_2 r_{2,j}(\hat{y}_j - x_{i,j}) \tag{1}$$

where w is the positive inertia weight, $v_{i,j}$ is the velocity of particle i in dimension j, $r_{1,j}$ and $r_{2,j}$ are uniquely generated random numbers in the interval $(0,1)$, c_1, c_2 are positive, cognitive and social constants, respectively. The position of each particle is updated according to the following equation:

$$x_{i,j} \leftarrow x_{i,j} + v_{i,j} \tag{2}$$

Given the above rules of position and velocity update, the particle swarm optimization algorithm can be expressed as follows:

```
Function $\hat{\mathbf{y}}$ = PSO ($\mathbf{x}_i$)
  If $\mathbf{x}_i$ == [] (empty matrix), initialize $\mathbf{x}_i$
  Initialize $\mathbf{v}_i$, $\mathbf{y}_i = \mathbf{x}_i$, $\hat{\mathbf{y}} = \arg\min_{\mathbf{x}_i} f(\mathbf{x}_i)$
  Repeat
    For each particle $i$
      Apply (1) to update velocity of each particle
      Apply (2) to update position of each particle
      Evaluate function $f(\cdot)$ at updated positions $\mathbf{x}_i$
      If $f(\mathbf{x}_i) < f(\mathbf{y}_i)$, update the local best values, $\mathbf{y}_i \leftarrow \mathbf{x}_i$
      If $f(\mathbf{y}_i) < f(\hat{\mathbf{y}})$, update the global best, $\hat{\mathbf{y}} \leftarrow \mathbf{y}_i$
  Until number of function evaluations < $max\_iter$
```

The fitness value of each particle is evaluated by a predefined observation model as follows:

$$f(\mathbf{x}_i) = p(\mathbf{o}_i|\mathbf{x}_i) \tag{3}$$

where $\mathbf{o}_i$ is the observation corresponding to $\mathbf{x}_i$.

The presented above particle swarm optimizer can be employed to carry out a global gradient-less stochastic search for the best configuration of the model parameters. The tracking of the human figure can also be formulated as the computation of the posterior probability distribution over the parameters of the model at time t given a sequence of images. Due to the nonlinearity of the likelihood function over model parameters the computation of the probability distribution is complicated. For these reasons the posterior is typically represented as a weighted set of particles, which are then propagated via a particle filter.

Particle filters approximate stochastically the state posterior with a set of N weighted particles, (s, π), where s is a sample state and π is its weight. This set of particles is propagated over time. At each time t the particles undergo selecting, predicting and re-weighting. In the select stage the algorithm randomly selects N particles from $\{s_{t-1}\}$ based on weights $\pi_{t-1}^{(n)}$. In the predict phase the particles undergo moving according to deterministic motion model. After the drift the particles are perturbed individually. Afterwards, based on observation model $p(\mathbf{o}_i|\mathbf{x}_i)$ the likelihood for each new sample is calculated, and finally weights are updated to obtain $\{s_t^{(n)}, \pi_t^{(n)}\}$.

In our approach to articulated motion tracking we exploit the power of the particle filter to represent multimodal distributions, which arise due to strong nonlinearity of the likelihood function over model parameters. Through the use of the particle filter less likely model configurations are not discarded immediately, but have chance to be considered in the next time. In articulated motion tracking the weakness of the particle filter consists in that the particles typically do not cluster around the true state of the figure and instead they concentrate around local maximas in the posterior distribution. In consequence, if particles are too diffused the tracking can be lost. In order to cope with this we employ particle swarm optimization in the particle filter to shift the particles toward more promising regions in the configuration space. The modified particle filter can be expressed as follows:

1. Select: Randomly select N particles from $\{s_{t-1}\}$ based on weights $\pi_{t-1}^{(n)}$.
2. Predict: Perturb individually particles
3. Shift: Shift the particles via the PSO
4. Re-weight: Get the likelihood for each new sample. Update the weights to obtain $\{s_t^{(n)}, \pi_t^{(n)}\}$
5. Estimate: Estimate the state using the mean $\mathbf{E}[\mathbf{x}_t] \approx \sum_{n=1}^{N} \pi_t^{(n)} s_t^{(n)}$

3 Experimental Results

The model of the human body has a form a kinematic chain consisting of 11 segments and the configuration of the model is defined by 26 DOF. The articulated model consists of cuboids modeling pelvis, torso, head, upper and lower arm and legs. A pose configuration is determined by position and orientation of the pelvis in the global coordinate system as well as relative angles between connected limbs. Given the parameters of the camera each cuboid can be projected into 2d image plane. To simplify the projection onto the image we project the corners via perspective projection and afterwards a rendering of the cuboids takes place. A regular rectangular grid is used to extract pixel values for each body part in such a rendered image.

Successful approaches to articulated object tracking typically rely on accurate extraction of foreground silhouettes using background subtraction. In [8][11][7] the edges have additionally been utilized in tracking of the human motion. In order to carry out qualitative analysis we construct foreground silhouettes through manual fitting of the 3d model to the person on the input images and then we render the model. This way we have in disposal the configuration of the human as well as the foreground image. Such a configuration reflecting the current human pose determines the reference image, which undergoes matching via the tracking algorithm. The tracking algorithm operates in 26 dimensional state space and generates pixel maps that are employed in computing the likelihoods.

The experiments were conducted on images acquired from surveillance cameras that are situated in a student hostel, see Fig. 1. For visualization purposes the reference sub-images were placed at the bottom left part of the input images. By projecting the body model into the images we can extract information for each body part as shown in the mentioned sub-images. The size of the input

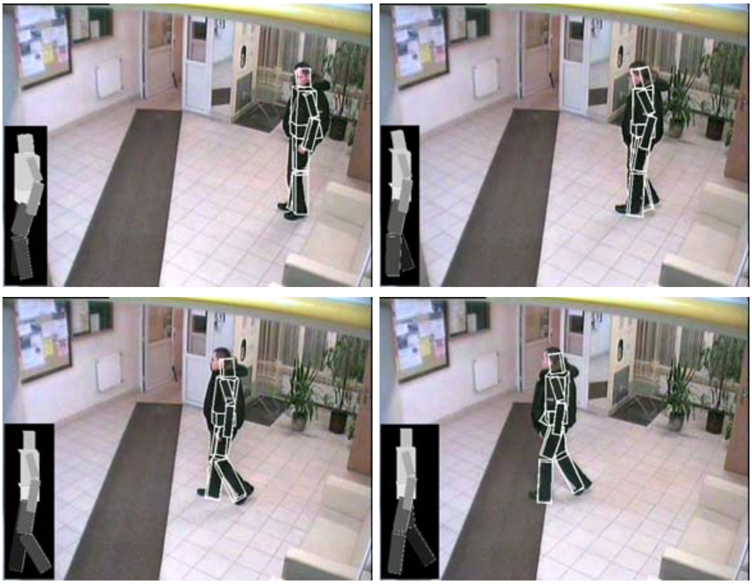


Fig. 1. 3d model-based human body tracking, frames #5, #15, #25 and #35, left bottom: appearance image of person undergoing tracking. The degree of overlap between the appearance image and the projected model into 2d image plane of the camera is 0.86, 0.80, 0.80 and 0.79, respectively.

images is 720x576 and they were acquired at 6 fps. The low frequency of the input sequence comprises considerable challenge for the examined algorithms.

A comparison among particle filter with embedded particle swarm optimization (PF+PSO), ordinary particle filter (PF), particle swarm optimization (PSO), kernel particle filter (KPF) [7], and annealed particle filter (APF) can be seen in Tab. 1. It can be observed that the PF+PSO algorithm is better in comparison to all remaining algorithms in term of the accuracy of body motion tracking. Both PSO+PF and PSO are superior to the remaining algorithms in terms of the computation time. Using 200 particles and 5 iterations in an unoptimized C/C++ implementation of the PF+PSO-based algorithm, a 2.0 GHz PC requires about 1.36 sec. per image to perform the motion tracking, most of the time being spent in the evaluation of the fitness function. In such a configuration of the tracker the estimates of the human pose in the sequence from Fig. 1 have acceptable accuracy. At the mentioned figure we demonstrate some experimental results, which were obtained using 500 particles and in 10 iterations. The results obtained via PSO are superior in comparison to results produced by KPF and APF. In the employed test sequence, which has been acquired with relatively low frequency,

Table 1. Computation time (4-th column) and average degrees of overlap between the reference image of human body and the estimated body pose (3-rd column) for particle filter (PF), particle swarm optimization (PSO), particle filter with particle swarm optimization (PF+PSO), kernel particle filter (KPF) and annealed particle filter (APF)

	#particles	#it.	overlap [%]	time [sec.]
PF	20000		0.75	22.92
	10000		0.74	11.39
	5000		0.73	5.69
	2000		0.67	2.71
PSO	1000	10	0.83	12.45
	500	10	0.80	6.20
	200	10	0.76	2.49
	1000	5	0.81	6.79
	500	5	0.78	3.40
	200	5	0.76	1.36
PF+PSO	1000	10	0.84	12.63
	500	10	0.83	6.24
	200	10	0.80	2.50
	1000	5	0.82	6.90
	500	5	0.81	3.41
	200	5	0.78	1.36
KPF	2000	3	0.76	7.26
	1000	3	0.74	3.50
APF	2000	10	0.79	22.38
	1000	10	0.78	11.16
	500	10	0.75	5.60

the KPF behaved better than APF. The discussed results are averages from three independent runs of the algorithms.

In Fig. 2 we demonstrate the degree of overlap versus frame number for the algorithms utilized in our experiments. It can be observed that the results obtained via PF+PSO algorithm are better. The PSO-based tracker is superior to PF+PSO tracker in the initial part of the image sequence. The magnitude of change of the overlap degree for PF+PSO is smaller in comparison to other curves. For PF+PSO the degree of overlap does not drop below 0.78, particularly in the end part of the sequence, where the remaining trackers achieve worse overlap degrees. The results for PF+PSO, PSO and APF were obtained using 500 particles and 10 iterations. The results for KPF were obtained with 2000 particles and 3 iterations, whereas 10000 particles were employed in PF. In Fig. 1 we can see how well the rendered model with configuration determined by PF+PSO fits the human silhouette, which has been shot in frames #5, #15, #25 and #35. In the experiments we have used simple motion models and we expect that prior model of the human motion can improve further the robustness as well as the accuracy of the motion tracking.

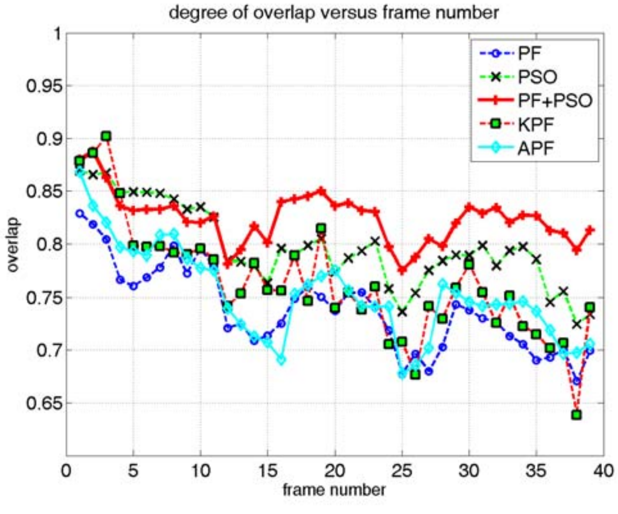


Fig. 2. Degree of overlap between the reference human body and the rendered image on the basis of the best particle

4 Conclusions

The most important general reason for the weakness of the particle filter is that in high dimensional state spaces the particles can not cluster around the true state and instead they have tendency to migrate towards local maximas in the posterior distribution. In this work, an effective algorithm for tracking human motion has been presented. The experimental results suggest that particle filter combined with particle swarm optimization achieves the best results in terms of the accuracy of the tracking. The algorithm is capable of tracking full articulated body motion efficiently. We have demonstrated the behavior of the algorithm on challenging human motion sequence. The algorithm has been compared with particle swarm optimization, kernel particle filter, annealed particle filter and ordinary particle filter. The tracker using particle swarm optimization achieves slight worse tracking accuracy than particle filter combined with particle swarm optimization. In image sequences acquired at low frame rates it outperforms both the annealed particle filter and the kernel particle filter.

Acknowledgment

This paper has been supported by the project "System with a library of modules for advanced analysis and an interactive synthesis of human motion" co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme - Priority Axis 1. Research and development of modern technologies, measure 1.3.1 Development projects.

References

1. Poppe, R.: Vision-based human motion analysis: an overview. Computer Vision and Image Understanding 108, 4–18 (2007)
2. Deutscher, J., Blake, A., Reid, I.: Articulated body motion capture by annealed particle filtering. In: IEEE Int. Conf. on Pattern Recognition, Hilton Head, South Carolina, USA, vol. 2, pp. 126–133 (2000)
3. Sidenbladh, H., Black, M., Fleet, D.: Stochastic tracking of 3d human figures using 2d image motion. In: European Conference on Computer Vision, pp. 702–718 (2000)
4. Isard, M., Blake, A.: Condensation - conditional density propagation for visual tracking. Int. J. of Computer Vision 29, 5–28 (1998)
5. Lawrence, N.D.: Gaussian process latent variable models for visualisation of high dimensional data. In: Advances in Neural Information Processing Systems (NIPS), Vancouver, Canada, pp. 329–336 (2003)
6. Sminchisescu, C., Triggs, B.: Covariance scaled sampling for monocular 3d body tracking. In: IEEE Int. Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 447–454 (2001)
7. Fritsch, J., Schmidt, J., Kwolek, B.: Kernel particle filter for real-time 3d body tracking in monocular color images. In: IEEE Int. Conf. on Face and Gesture Rec., Southampton, UK, pp. 567–572. IEEE Computer Society Press, Los Alamitos (2006)
8. Balan, A., Sigal, L., Black, M.: A quantitative evaluation of video-based 3d person tracking. In: IEEE Workshop on VS-PETS, pp. 349–356 (2005)
9. Ivekovic, S., Trucco, E., Petillot, Y.R.: Human body pose estimation with particle swarm optimisation. Evolutionary Computation 16, 509–528 (2008)
10. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proc. of IEEE Int. Conf. on Neural Networks, pp. 1942–1948. IEEE Press, Piscataway (1995)
11. Deutscher, J., Reid, I.: Articulated body motion capture by stochastic search. Int. J. Comput. Vision 61, 185–205 (2005)

GPU-Accelerated Tracking of the Motion of 3D Articulated Figure

Tomasz Krzeszowski, Bogdan Kwolek, and Konrad Wojciechowski

Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008 Warszawa
http://www.pjwstk.edu.pl

Abstract. This paper presents methods that utilize the advantages of modern graphics card hardware for real-time full body tracking with a 3D body model. By means of the presented methods the tracking of full body can be performed at frame-rates of 5 frames per second using a single low-cost moderately-priced graphics card and images from single camera. For a model with 26 DOF we achieved 15 times speed-up. The pose configuration is given by the position and orientation of the pelvis as well as relative joint angles between the connected limbs. The tracking is done through searching for a model configuration that best corresponds to the observed human silhouette in the input image. The searching is done via particle swarm optimization, where each particle corresponds to some hypothesized set of model parameters.

1 Introduction

The era of increase of performance of single-threaded programs at single silicon chip has come to an end. Programs will only increase the performance if they utilize parallelism. Recently, leading GPU vendors make available GPU programming environments. For instance, NVIDIA introduced CUDA environment to perform data-parallel calculations on GPU. As a result, there have been reported several studies in the literature that exploited GPUs for accelerating algorithms, including image processing and recognition algorithms. GPUs provide the best cost-per-performance parallel architecture for data-level parallelism with high computing demands. The performance bottleneck of most implementations intended for execution on GPU is memory access. Therefore, the algorithms to be executed on GPU should be carefully designed in order to achieve good memory performance, which leads to considerable speed-up of the computations. Thus, GPUs are not the best choice for all computer vision problems.

Non intrusive human body tracking is a key issue in advanced human-computer communication. This is one of the most challenging problems in computer vision being simultaneously one of the most computationally demanding tasks. For example, a tracker [1] employing 10 annealing layers with 200 particles needed around 1 hour to process 5 seconds of footage. Considerable amount of work has been done to achieve reliable and fast articulated motion tracking [2][1][3][4][5]. However, to the best of our knowledge, no GPU implementation of articulated body tracking has been developed until now.

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 155–162, 2010.

2 Programming of GPU

In this Section we discuss the architectural features of G80, which are most relevant to understand our implementation. The G80 graphics processing unit architecture was first introduced in NVIDIA's GeForce 8800 GTS and GTX graphics cards. A GTX 280 card that is compatible with G80 and supports Computing Capability 1.3 has been used in our experiments. It has 240 cores in 30 streaming 1.3 GHz multiprocessors, which support Single Program Multiple Data (SPMD) programming model.

The programming of GPU has been considerably simplified through introducing CUDA framework by NVIDIA. CUDA makes programming of GPU easier as it hides hardware details allowing a programmer to think in terms of memory and arithmetic operations, rather than in categories of primitives and textures being specific to graphical operations. To obtain the best performance from G80-based GPUs, we have to keep all processors occupied and hide memory latency. In order to achieve this aim, CUDA supports running hundred or thousands of lightweight threads in parallel. No extra code is needed for thread management, and the CPU is capable of running concurrently with the GPU. In CUDA the programs are expressed as kernels. A part of the application that operates on different elements of a dataset can be isolated into a kernel that is executed on the GPU by many different threads. Kernels run on a grid, which is an array of blocks, whereas each block is an array of threads. Blocks are mapped to multiprocessors and each thread is mapped to a single core. Threads within a block are grouped into warps.

At any time a multiprocessor can execute a single warp. Every thread of a warp executes the same instruction but operates on different data. A unique set of indices is assigned to each thread to determine to which block it belongs and its location inside it. Threads in one block can communicate each other using the shared memory, but two threads from two different blocks cannot cooperate via shared memory. The GPU handles latency by supporting thousands of threads in flight at once. In current GPUs, context switch is very fast because everything is stored in registers and thus there is almost no data movement. The card's DRAM memory is accessible from different blocks. It is, however, much slower than the on-chip shared memory. Its latency can be hidden by careful design of control flow as well as design of kernels. To achieve good performance both high density of arithmetic instructions per memory access as well as several hundreds of threads per block are needed. This permits the GPU to execute arithmetic instructions while certain threads are waiting for access to the global memory.

3 Parallel PSO for Object Tracking

Particle swarm optimization (PSO) [6] is a global optimization, population-based evolutionary algorithm for dealing with problems in which a best solution can be represented as a point in a n-dimensional space. The PSO is initialized with a group of random particles (hypothetical solutions) and then it searches hyperspace (i.e. R^n) of a problem for optima. Particles move through the solution space, and undergo evaluation according to some fitness function after each time step. The particles iteratively evaluate their candidate solutions and remember the location of their best location with the smallest objective value so far, making this information available to their neighbors. Particles communicate good positions to each other and adjust their own velocity and then the position based on such good positions. Additionally each particle employs a best value, which can be:

- a global best that is immediately updated when a new best position is found by any particle in the swarm
- neighborhood best where only a specific number of particles is affected if a new best position is found by any particle in the sub-population

A topology with the global best converges faster as all the particles are attracted simultaneously to the best part of the search space. Neighborhood best allows parallel exploration of the search space and decreases the susceptibility of falling into local minima, however, it slows down the convergence speed. Taking into account the computational overheads the topology with global best is utilized in our approach.

In the ordinary PSO algorithm the update of particle velocity and position is given by the following equations:

$$v_j^{(i)} \leftarrow wv_j^{(i)} + c_1 r_{1,j}^{(i)}(p_j^{(i)} - x_j^{(i)}) + c_2 r_{2,j}^{(i)}(p_{\mathrm{g},j} - x_j^{(i)}) \tag{1}$$

$$x_j^{(i)} \leftarrow x_j^{(i)} + v_j^{(i)} \tag{2}$$

where w is the positive inertia weight, $v_j^{(i)}$ is the velocity of particle i in dimension j, $r_{1,j}^{(i)}$ and $r_{2,j}^{(i)}$ are uniquely generated random numbers with the uniform distribution in the interval [0.0, 1.0], c_1, c_2 are positive constants, $p^{(i)}$ is the best position that the particle i has found, p_{g} denotes best position that is found by any particle in the swarm.

The velocity update equation (1) has three main components. The first component, which is often referred to as inertia models the particle's tendency to continue the moving in the same direction. In effect it controls the exploration of the search space. The second component, called cognitive, attracts towards the best position $p^{(i)}$ previously found by the particle. The last component is referred to as social and attracts towards the best position p_{g} found by any particle. The fitness value that corresponds $p^{(i)}$ is called local best $p_{\mathrm{best}}^{(i)}$, whereas the fitness value corresponding to p_{g} is referred to as g_{best}.

Given the above equations the PSO algorithm can be illustrated in the following manner:

1. Assign each particle a random position in the problem hyperspace.
2. Evaluate the fitness function for each particle.
3. For each particle i compare the particle's fitness value with its $p_{\mathrm{best}}^{(i)}$. If the current value is better than the value $p_{\mathrm{best}}^{(i)}$, then set this value as the $p_{\mathrm{best}}^{(i)}$ and the current particle's position $x^{(i)}$ as $p^{(i)}$.

4. Find the particle that has the best fitness value g_{best}.
5. Update the velocities and positions of all particles according to (1) and (2).
6. Repeat steps $2-5$ until a stopping criterion is not satisfied (e.g. maximum number of iterations or a sufficiently good fitness value is not attained).

Our parallel PSO algorithm for object tracking consists of five main phases, namely initialization, evaluation, *p_best*, *g_best* and update. At the beginning of each frame, in the initialization stage an initial position $x^{(i)} \leftarrow \mathcal{N}(p_{\text{g}}, \Sigma)$ is assigned to each particle, given the location p_{g} that has been estimated in the previous frame. In the evaluation phase the fitness value of each particle is calculated using a predefined observation model as follows:

$$f(x^{(i)}) = p(o^{(i)}|x^{(i)}) \tag{3}$$

where $o^{(i)}$ is the observation corresponding to $x^{(i)}$. It is the most time consuming operation on GPU. The calculation of the observation model is discussed in Section 4.2 and the decomposition of this operation into kernels is presented in Section 4.3. In the *p_best* stage the determining of $p^{(i)}_{\text{best}}$ as well as $p^{(i)}$ takes place. This stage corresponds to operations from the point 3. of the presented above PSO pseudo-code. The operations mentioned above are computed in parallel using available GPU resources, see Fig. 1. Afterwards, the g_{best} and its corresponding p_{g} are calculated in a sequential task. Finally, the update stage that corresponds to point 5. in the PSO pseudo-code is done in parallel. That means that in our implementation we employ the parallel synchronous particle swarm optimization. The synchronous PSO algorithm updates all particle velocities and positions at the end of every optimization iteration. In contrast to synchronous PSO the asynchronous algorithm updates particle positions and velocities continuously using currently accessible information.

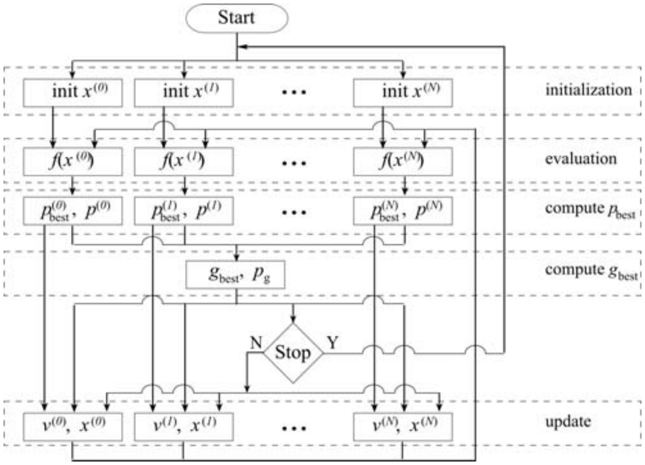

Start
init $x^{(0)}$
init $x^{(1)}$
init $x^{(N)}$
initialization
$f(x^{(0)})$
$f(x^{(1)})$
$f(x^{(N)})$
evaluation
$p^{(0)}_{\text{best}}, p^{(0)}$
$p^{(1)}_{\text{best}}, p^{(1)}$
$p^{(N)}_{\text{best}}, p^{(N)}$
compute p_{best}
$g_{\text{best}}, p_{\text{g}}$
compute g_{best}
N
Stop
Y
$v^{(0)}, x^{(0)}$
$v^{(1)}, x^{(1)}$
$v^{(N)}, x^{(N)}$
update

Fig. 1. Decomposition of synchronous particle swarm optimization algorithm on GPU

4 Implementation of Articulated Body Tracking on GPU

4.1 Tracking Framework

The articulated model of the human body has a form a kinematic chain consisting of 11 segments. The 3d model is constructed using cuboids that model the pelvis, torso, head, upper and lower arm and legs. The configuration of the model is defined by 26 DOF. It is determined by position and orientation of the pelvis in the global coordinate system and the relative angles between connected limbs. Each cuboid can be projected into 2d image plane via perspective projection. To obtain a projected image of the model we transform the corners via perspective projection and afterwards we perform a rendering of the cuboids. This way we obtain the image of the 3d model in a given configuration. During calculation of the fitness function we employ a regular rectangular grid to extract pixel values for each body part in such a rendered image.

In most of the approaches to articulated object tracking a background subtraction algorithms are employed to extract a person undergoing tracking. Additionally, image cues such as edges, ridges, color are often employed to improve the extraction of the person. In this work the images of the person undergoing tracking are synthesized on the basis of 3d models of the human body. The parameters are determined using model in a configuration, which has been manually determined through fitting of the model to the person on the input images. Given such parameters we generate synthesized images of the human body. In the experiments we employ animations generated via Blender application[1] as well as images that are generated on the basis of the same 3d model that is used in tracking, see sub-images in Fig. 2.

4.2 Observation Model

The most computationally and time demanding operation is generation of the body image on the basis of the hypothesized body configurations established by particles. Precisely speaking, the most computationally intensive operation is the rasterization of the triangles. The rendering stage creates a two dimensional display of triangles given the transformed vertexes of the 3d model. It involves the calculation of the pixels forming the triangles. GPU designers have incorporated many rasterizatiom algorithms over the years. In all rasterization algorithms the pixel is treated independently from all other pixels. Therefore, the GPU can handle all pixels in parallel.

The so-called painter's algorithm consists in sorting the object or polygons from back to front and then rasterizing them in that order. Currently, a modified painter's algorithm is used to perform the depth test. In the parallel rendering algorithm, which is based on a modified version of the painter's algorithm, we perform painting in the reverse order, i.e. we first paint out the nearest element. Afterwards, we paint out the triangles according to the order of the model parts.

[1] http://www.blender.org/

In order to paint out a given triangle we determine the surrounding rectangular sub-image and then we verify all pixels of such a sub-image. If a considered pixel had not been previously painted out we verify if it belongs to the considered triangle. If yes, we paint it out.

The fitness function (3) is determined on the basis of the overlap degree between the reference image of the body and the current rasterized image. The overlap degree is calculated through checking the overlap from the reference to the current rasterized image as well as from the current rasterized image to the reference body. The larger the degree overlap is, the larger is the fitness value. Figure 2 depicts some images used in the experiments. The images were acquired by surveillance cameras in a student hostel. In the sub-images at the bottom-left the reference images are shown.

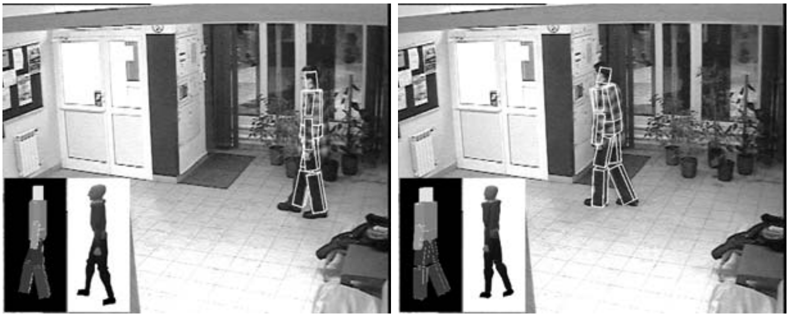


Fig. 2. 3d model-based human body tracking, frames #5, #40, left bottom: appearance images of person undergoing tracking. The overlap degree between the appearance image and the projected model into 2d image plane is 0.84 and 0.86, respectively.

4.3 Algorithm Decomposition

In order to decompose an algorithm into GPU we should identify data-parallel portions of the program and isolate them as CUDA kernels. In the initialization stage we generate pseudo-random numbers using the Mersenne Twister [7] kernel provided by the CUDA^{TM} SDK. From uniform random numbers we generate a vector of normal random numbers using Box Mueller transform based on trigonometric functions [8] to initialize the positions of the particles. At the beginning of each iteration we generate the random numbers for all particles through single call of the kernel. Taking into account that the maximum number of threads in one block is 512, in one block we initialize 19 particles. In the evaluation phase we employ two kernels. The first kernel is used in rendering of the 3d body model into 2d plane, whereas the second one in calculation of the measure similarities between projections of the 3D model and the content of the reference images. In our approach each block is responsible for rendering one image. Taking into account the available number of registers we run 448 threads and each thread is in charge of painting out of several pixels. In order to obtain the degree of overlap the comparison of the images is done using one dimensional textures and a single thread compares the pixels from two corresponding image columns. In the update phase each thread is responsible for updating one dimension of the particle's location.

5 Experiments

In this Section, we first compare the runtimes of our GPU and CPU implementations and present our speedup. Then, we show the tracking performance using synchronous and asynchronous implementations of PSO. This is followed by a discussion of the factors that limit our performance.

The experiments were conducted on a notebook with 4 GB RAM, Intel Core 2 Duo, 2 GHz processor with GT 130M graphics card. The graphics card has 4 stream multiprocessors with 1.5 GHz, each with 8 cores. It is equipped with 512 MB RAM, 64 KB constant memory and 16 KB common memory. We conducted also experiments on a PC with single NVIDIA GTX 280 card. The card has 30 stream multiprocessors with 1.3 GHz clock, each with 8 cores. It has 1 GB RAM, 64 KB constant memory and 16 KB common memory.

Table 1 shows computation time that has been obtained on CPU, GT 130M and GTX 280. Using the PSO algorithm with 500 particles and 5 iterations we can process in real-time 5 frames per second. The average degree of overlap between the reference body image and the projected body with the estimated configuration in the 50 frame long sequence is slightly below 0.8. The results in table demonstrate that the mobile graphics card was also capable of obtaining a speed-up.

Table 1. Computation time [sec.]

	CPU	GT 130M	GTX 280
#4000, 10 it.	48.89	20.35	2.94
#2000, 10 it.	24.51	10.06	1.49
#1000, 10 it.	12.28	5.26	0.75
#500, 10 it.	6.12	2.65	0.39
#4000, 5 it.	26.74	11.19	1.59
#2000, 5 it.	13.38	5.52	0.81
#1000, 5 it.	6.68	2.87	0.41
#500, 5 it.	3.34	1.45	0.22

We compared the effectiveness of the synchronous and asynchronous version of the PSO algorithm. The asynchronous PSO that is used in our CPU implementation gives something better results. For instance, for a set-up with 2000 particles and 10 iterations the overlap degree for asynchronous PSO is equal to 0.85, whereas for synchronous version it is equal to 0.80. In a set-up with 500 particles and 10 iterations the overlap degree is equal to 0.79 and 0.78, respectively.

The most time-consuming operation of the tracking algorithm is the rendering of the 3d model. This operation amounts to 0.92 of whole processing time. The comparing of images in order to determine the degree of overlap amounts to 0.05 of full amount of processing time.

6 Conclusions

In this paper we presented an algorithm for articulated human motion tracking on GPU. The articulated model of the human body consists of 11 segments and has 26 DOF. We showed that our GPU implementation has achieved a speedup of more fifteen times than our CPU-based implementation. The tracking of full body can be performed at frame-rates of 5 frames per second using a single low-cost graphics card and single camera images. With rapid development of the graphics card technologies, the tracking speed is expected to be further accelerated in the near future by newer generations of the GPU architecture.

Acknowledgment

This paper has been supported by the project "System with a library of modules for advanced analysis and an interactive synthesis of human motion" co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme - Priority Axis 1. Research and development of modern technologies, measure 1.3.1 Development projects.

References

1. Deutscher, J., Blake, A., Reid, I.: Articulated body motion capture by annealed particle filtering. In: IEEE Int. Conf. on Pattern Recognition, pp. 126–133 (2000)
2. Poppe, R.: Vision-based human motion analysis: an overview. Computer Vision and Image Understanding 108, 4–18 (2007)
3. Fritsch, J., Schmidt, J., Kwolek, B.: Kernel particle filter for real-time 3D body tracking in monocular color images. In: IEEE Int. Conf. on Face and Gesture Rec., Southampton, UK, pp. 567–572. IEEE Computer Society Press, Los Alamitos (2006)
4. Zhao, T., Nevatia, R., Wu, B.: Segmentation and tracking of multiple humans in crowded environments. PAMI 30, 1198–1211 (2008)
5. Wu, C., Aghajan, H.K.: Human pose estimation in vision networks via distributed local processing and nonparametric belief propagation. In: Blanc-Talon, J., Bourennane, S., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2008. LNCS, vol. 5259, pp. 1006–1017. Springer, Heidelberg (2008)
6. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proc. of IEEE Int. Conf. on Neural Networks, pp. 1942–1948. IEEE Press, Piscataway (1995)
7. Matsumoto, M., Nishimura, T.: Mersenne twister: a 623-dimensionally equidistributed uniform pseudorandom number generator. ACM Trans. Model. Comput. Simul. 8, 3–30 (1998)
8. Box, G.E.P., Muller, M.E.: A note on the generation of random normal deviates. The Annals of Mathematical Statistics 29, 610–611 (1958)

An Efficient Approach for Human Motion Data Mining Based on Curves Matching

Van-Hanh Nguyen[1], Frederic Merienne[2], and Jean-Luc Martinez[3]

Arts et Métier ParisTech - Le2i - CNRS
3 Rue Thomas Dumorey, 71100 Chalon Sur Saone, France
hanhnv@gmail.com, frederic.Merienne@ensam.eu,
jean-luc.martinez@cluny.ensam.eu

Abstract. In this paper, we present a novel and efficient approach to retrieve human motion capture data as used in data-driven computer games, animated movies and special effects in the aim of finding a specific motion. From the kinematic chain model, the human motion capture data is transformed to a spatial-temporal invariance representation called the motion feature representation, in which each segment of kinematic chain model is represented by an angle between itself and the root segment. We treat the human motion as a cluster of curves of angle. In the aim of finding a human motion capture data in a very large database, we propose a novel lower bounding distance called LB_Keogh_Lowe to speed up similarity search. In order to reduce the computational cost, we employ techniques to simplify the curves length of both the envelopes curves and the query data. The similarity between two human motions is measured by applying the constrained Dynamic Time Warping. We carry out an experimental analysis with various real motion capture dataset. The results demonstrate the efficiency of our approach in the context of the human motion capture data and the potentiality to apply it in others contexts of the time-series data retrieval.

1 Introduction

Nowadays, motion capturing plays an important role in computer games, computer-animated movies or video games as a tool for synthesizing realistic motion sequences. In general, by using optical devices, trajectories of moving markers attached to the human body are recorded as 3D motion data. The huge quantity of motion capture dataset is very expensive and time consuming due to the high cost of recording equipment, high computational power and the requirement of manual post-processing operations. Therefore, it has been suggested reusing the motion capture data by modifying and adapting the existing motion clips via editing and morphing techniques [1][2][3]. Hence, there is a strong need to develop an efficient method to retrieve a motion data among others.

In practice, content-based approach is one of the most efficient methods to enable an automatic extraction of a user-specified motion. The crucial point of this approach is to define the feature of motion content utilized to qualify the

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 163–184, 2010.

motion. Some existing methods consider the motion key-frame as the feature but cannot handle the spatial-temporal variation of motion [4]. In this paper, we propose a novel approach consisting in transforming the motion capture data to a spatial invariance space called motion feature representation. For that, we use the kinematic chain model to represent the motion capture date. In the spatial invariance space, each segment of kinematic chain model is represented by an angle between itself and the root segment. Thus, we treat the transformed motion as a cluster of curves of all the angles describing the kinematic motion. The distance between two motions is defined as the similarity between the curves of the motions. Hence, we conduct the motion data mining as the shape matching problem between the curves. In the aim of finding a human motion capture data in a very large database, we propose techniques to speed up similarity search based on bounding envelopes curves. In order to reduce the computational cost, we employ techniques to simplify the curves length of both the envelopes curves and the query data. The similarity between two human motions is measured by applying the constrained Dynamic Time Warping technique on the curves of their angles. Our contributions can be summarized as following:

- We introduce the spatial-temporal invariance representation for human motion capture data and conduct the human motion data mining to the curves matching problem.
- We present a novel lower bounding distance called LB_Keogh_Lowe for pruning power the retrieval of human motion capture data.
- We demonstrate the efficiency and the effectiveness of our method on the various real human motion datasets.

The rest of the paper is organized as follows. In the next section, we review some discussion of related work. In Section 3, we describe the necessary geometric formulas which are utilized in our human motion data mining approach. Section 4 and Section 5 focus on our approach of human motion data mining technique. Section 4 provides an overall view of the proposed human motion retrieval system which includes the chosen human motion features, the technique to measure the similarity between two motions and the low bounding technique to speed up the searching. Section 5 investigates some proposed techniques for more pruning power the retrieval of human motion capture data. Experiments and results discussion are employed in Section 6. We finish with the conclusion and further work in Section 7.

2 Related Work

In computer animation or movie specified effects, data-driven motion synthesis is an important technique to generate the realistic motions from the recorded motion capture data. Due to the large use of motion capture data, the efficient reuse of such data is more and more challenging. In fact, human motion capture data is a kind of high-dimensional time-series and its raw data describe the spatial information of a large number of markers in the three-dimensional space, thus the problem of human motion data mining is considered as the time-series retrieval. In terms of time-series matching, three main aspects are considered. The first one is how to represent the human motion data. The second one is how to determine the distance function to measure the similarity of two motions and the last one concerns the considered techniques to speed up the searching.

In order to deal with motion capture data, several representations have been proposed. Lee et al. [6] described a two-layer structure for representing human motion capture data, but motion physical features cannot be represented clearly in this case. Chui et al. [7] proposed local spherical coordinates relative to the root orientation as the segments posture of each skeletal segment. But the skeletal segment is represented by two parameters which cannot be used to observe posture of each skeletal segment. In our work, we utilize a spatial invariance representation from the work of Xiao et al. [5]. The representation is based on the bone angle and is effective for human motion retrieval. Opposed to the work of Xiao et al. [5] which considers the human motion as frames, we treat the human motion as a curves cluster of the bone angles, thus the problem of temporal variation could be handled.

In fact, the human motion data mining is a part of time-series data mining in a very large database (VLDB) domain. In terms of the distance function to measure the similarity of the time-series data, there are many proposed methods classified into two main approaches: Model-based approach such as Hidden Markov Model (HMM)[8], Neural Network (NN)[9], and Distance function-based approach, which is differentiated into two sub-approaches, in which the first one is called the Metric functions such as Euclidean distance, Manhattan distance or Normal distance [10] and the second one is called the Non-metric functions such as constrained Dynamic Time Wrapping (cDTW)[14], wLCSS[17], Edit distance with Real Penalty (ERP)[11], Edit Distance on Real sequence (EDR)[12], Spatial Assembling Distance (SpADe)[15], etc. In that domain, the experimental results of Keogh et al., 2002 [14] have to be mentioned. Keogh et al. showed that the performance of the distance function-based approach is much better than that of the model-based approach, for both criteria of accuracy and computational cost. Moreover, the work of Ding et al., 2008 [15] performed a comparison of the major techniques of distance function-based approach by testing their effectiveness on 38 time series data set from a wide variety of application domains. Experimental results obtained show that, in over a dozen of distance measures of similarity among time series data in the literature, cDTW [14][16] is slightly better than the others from accuracy point of view. Moreover, on large data set, computation cost of cDTW is very close to Euclidean distance algorithm. Indeed, cDTW is our chosen distance function to measure the similarity of the human motion capture data.

In terms of speed up the searching, there are two techniques utilized. The first one consists in reducing the length of the human motion by using approximating representations. The other one consists in quickly filtering the negative time-series due to the fact that the number of the positive time-series which match with the candidature sequence is less much than the number of the negative ones. Many techniques have been proposed in the literature for representing time series with reduced dimensionality. Most of them have been reviewed in work of Ding et al. [15], such as Discrete Fourier Transformation (DFT) [18], Single Value Decomposition (SVD) [18], Discrete Cosine Transformation (DCT) [19], Discrete Wavelet Transformation (DWT) [20], Piecewise Aggregate Approximation (PAA) [22], Adaptive Piecewise Constant Approximation (APCA) [21], Chebyshev polynomials (CHEB) [23], Symbolic Aggregate approXima-tion (SAX) [24], Indexable Piecewise Linear Approximation (IPLA) [25] and etc. In our work, we approach another approximating representation called curve simplification representation which is proposed by Lowe [26]. The experimental result shows the efficiency and effectiveness of that approach in the human motion database [27][28]. Opposed to the original version of the curve simplification algorithm, we propose modified versions adapting with the human motion data mining problem. To speed up the searching by filtering rapidly the negative time-series sequences, lower bounding is popular technique utilized. The main purpose of the using the lower bounding function is due to the fact that, the computational complexity of the measurement distance is in the order of *O(n2)* in general while using lower bounding technique we are capable of pre-filtering the negative sequences with the complexity in the order of *O(n).* Thus, we should avoid using the measurement distance for the negative sequences. Moreover, the number of the negative sequences is much more than the number of potential positive ones in the real datasets. Actually, due to the utilized measurement distance of $cDTW$ in the paper, there are several lower bounding functions available for $cDTW$ such as the lower bounding function introduced by Kim et al. [29] or another one proposed by Yi et al. [30]. In addition, Keogh et al. [14] proposed a more efficient lower bounding function referred as LB_Keogh. Their experimental results demonstrate the effectiveness of LB_Keogh in terms of time-series data retrieval. Indeed, our technique is based on the LB_Keogh which is described in Section 4 of the paper.

3 Geometric Formulas for the Human Motion Data Mining Approach Based on Curves Matching

In this section, we describe the necessary geometric formulas which are utilized during the paper. In prior, we describe some definitions relative to our work.

- **Coordinate system of the curve:** The coordinate system of the curve is the 2D space which is defined by vertical axis illustrated in degree unit and horizontal axis illustrated in frame number unit (temporal axis of the human motion).
- **Time-series curve:** In our work the curve of m time-series points is referred as time-series curve which is denoted as follows:

$$C(m) = \{c_i(p_{c,i}, t_{c,i})\}, i = 1..m; t_i > \ t_j \ \forall i > j \tag{1}$$

where $t_{c,i}$ is measured in the order of the frame number unit and $p_{c,i}$ is measured in degree unit and m is the length or the dimension of the curve.

Due to the fact that $t_{c,i} > t_{c,j} \, \forall i > j$, thus the time-series curves are the forward curves or the none-self-intersecting curves. In the paper, we compare curves which have the same frame number value of the beginning point and the end point, that means the $t_{c,1}(t_{c,n})$ of any curve refers the same value. Having defined the necessary terms, we conduct to describe the geometric formulas utilized in our work.

3.1 Area of the Polygon Constituted by Two Time-Series Curves

The first formula in our work concerns the polygon area composed by the two time-series curves U(m) and V(n) to compare. We differentiate two cases. The first case is composed by all configurations of U(m) and V(n) with no intersecting points between them. Then, the polygon is the none-self-intersecting polygon which is illustrated in Figure 1. We denote S_1 as the area of the polygon. Formula 2 is described as below:

$$S_1(U,V) = \left\| \frac{s_1 + s_2 + s_3 + s_4}{2} \right\| \tag{2}$$

where $s_1 = \sum_{i=1}^{m-1} sign(u_i) \times (p_{u,i} \times t_{u,i+1} - p_{u,i+1} \times t_{u,i})$, $s_2 = sign(u_m) \times (p_{u,m} \times t_{v,n} - p_{v,n} \times t_{u,m})$, $s_3 = sign(v_1) \times (p_{u,1} \times t_{v,1} - p_{v,1} \times t_{u,1}$, and $s_4 = \sum_{j=2}^{n} sign(v_j) \times (p_{v,j-1} \times t_{v,j} - p_{v,j} \times t_{v,j-1})$ and $sign(u_i)$ or $sign(v_j$ is the sign function which obtains the value of 1 or -1 denoted the sign of the polygon edges illustrated in Figure 1 in which, $sign(u_i), \forall i = 1...m-1$ denotes the sign of the edge made by the point ui and u_{i+1}, sign(um) denotes the sign of the edge made by the point um and the point vn, $sign(v_j), \forall j = 2...n$denotes the sign of the edge made by the point v_{j-1} and the point v_j, and $sign(v_1)$ denotes the sign of the edge made by the point u_1 and the point v_1.

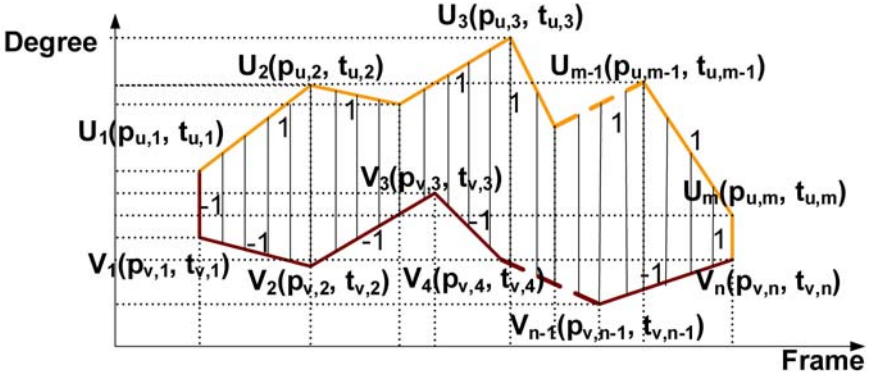


Fig. 1. Area of the none-self-intersecting polygon (drawn by the vertical lines) constituted by two time-series curves

In fact, we obtained the formula as above due to the familiar formula to calculate the area of the none-self-intersecting planar polygon described that: *The area of a parallelogram and triangle can be expressed as the magnitude of the cross-product of two edge vectors* [33]. In the case of non-self-intersecting polygon we could assume that $sign(u_i) = 1$ and the $sign(v_j) = -1, \forall i, j$ thus s_1, s_2, s_3, s_4 could be reduced as :$s_1 = \sum_{i=1}^{m-1}(p_{u,i} \times t_{u,i+1} - p_{u,i+1} \times t_{u,i})$, $s_2 = p_{u,m} \times t_{v,n} - p_{v,n} \times t_{u,m}$, $s_3 = p_{v,1} \times t_{u,1} - p_{u,1} \times t_{v,1}$, and $s_4 = \sum_{j=2}^{n}(p_{v,j} \times t_{v,j-1} - p_{v,j-1} \times t_{v,j})$

In practice, a technique with the computational complexity in the order of *O(n)* can be performed to calculate the area of the none-self-intersecting polygon.

The second case is illustrated in Figure 2, in which two time-series curves may have intersecting points. We refer the polygon of this case as the self-intersecting polygon. Recall the formula 2, we denote S_2 as the area of the self-intersecting polygon thus S_2 is formulated as in Formula 3.

$$S_2(U, V) = \left\| \frac{s_1 + s_2 + s_3 + s_4}{2} \right\| \tag{3}$$

where $s_1 = \sum_{i=1}^{m-1} sign(u_i) \times (p_{u,i} \times t_{u,i+1} - p_{u,i+1} \times t_{u,i})$, $s_2 = sign(u_m) \times (p_{u,m} \times t_{v,n} - p_{v,n} \times t_{u,m})$, $s_3 = sign(v_1) \times (p_{u,1} \times t_{v,1} - p_{v,1} \times t_{u,1}$, and $s_4 = \sum_{j=2}^{n} sign(v_j) \times (p_{v,j-1} \times t_{v,j} - p_{v,j} \times t_{v,j-1})$.

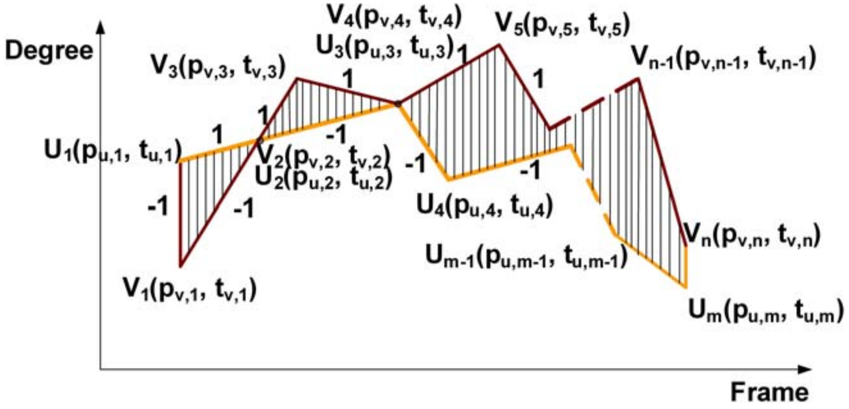
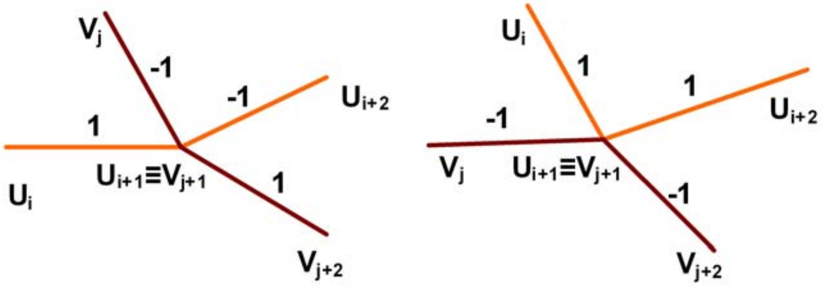


Fig. 2. Area of the self-intersecting polygon (drawn by the vertical lines) constituted by two time-series curves

At first glance, Formula 3 seems to be the same as Formula 2, but in reality the self-intersecting polygon constituted by two curves is added the intersecting points between them, such as the intersecting point $v_2 \equiv u_2$ or the intersecting point $v_4 \equiv u_3$ as illustrated in Figure 2. In this case, the sign function of the edges will be determined based on the sign of the previous edge and the convenience of equalities described in Figure 3. In left visual diagram, the signs of edges

(u_{i+1}, u_{i+2}) or (v_{j+1}, v_{j+2}) change due to the signs of their previous edges. In the right one, the signs are not changing.

In practice, intersecting points are unknown when occurring between two curves, so we should have a function to sort out the intersecting points between two edges. Anyways, a technique with the computational complexity in the order of *O(n)* could be considered to calculate the area of the self-intersecting polygon.



If $(p_{u,i} - p_{v,j}) \times (p_{u,i+2} - p_{v,j+2}) < 0$ the sign of the segment (U_{i+1}, U_{i+2}) and the segment (V_{i+1}, V_{i+2}) change from their privious segments

If $(p_{u,i} - p_{v,j}) \times (p_{u,i+2} - p_{v,j+2}) \geq 0$ the sign of the segment (U_{i+1}, U_{i+2}) and the segment (V_{i+1}, V_{i+2}) are maintained from their privious segments

Fig. 3. Two considered cases to determine the changing of the sign functions of the edge from the its previous edge

3.2 Area of the Outlier Parts Generated by the Time-Series Curve and the None-Self-intersecting Polygon Constituted by Two Other Time-Series Curves

Having the formulas to calculate the area of both the none-self-intersecting polygon and the self-intersecting polygon, we calculate the area of the outlier parts which are generated by a time-series curve denoted C(l) and a none-self-intersecting polygon constituted by two others time-series curves noted U(m) and V(n). As we described above, the three curves have the same frame number value of the beginning point and the end point. Figure 4 is an example illustrating the formula in which, the outlier parts are drawn by the vertical line which consists of three sub-polygons s_2, s_5 and s_7. S(C,U,V) is the area of the outlier parts. S1(U, V) is the area of the none-self-intersecting polygon constituted by U(m) and V(n). S2(U, C) and S2(C, V) are noted the areas of the self-intersecting polygon constituted by U(m) and C(l), and the self-intersecting polygon constituted by C(l) and V(n), respectively. The formula to calculate S(C,U,V) is described as below in Formula 4.

$$S(C, U, V) = \frac{S_2(U, C) + S_2(C, V) - S_1(U, V)}{2} \tag{4}$$

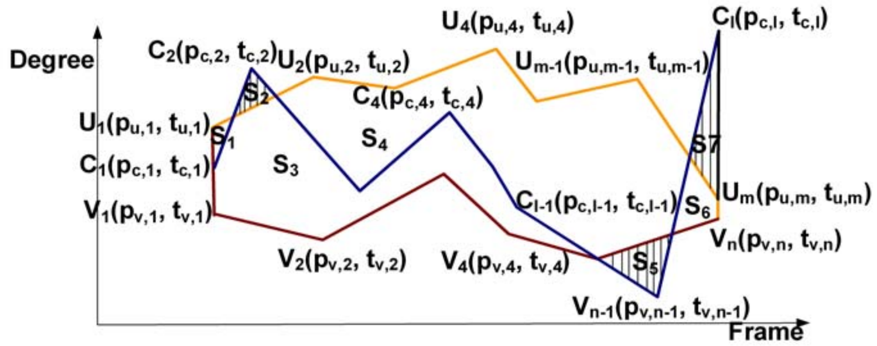


Fig. 4. Area of the outlier parts between a curve C(l) and a none-self-intersecting polygon constituted by two other curves U(m) and V(n). The outlier parts are drawn by the vertical lines. Denote S(C,U,V) as the area of those parts, si is the area of the ith sub-polygon generated by the curves illustrated as in Figure 4. We have: $S(C, U, V) = s_2 + s_5 + s_7; S_1(U, V) = s_1 + s_3 + s_4 + s_6; S_2(U, C) = s_1 + s_2 + s_4 + s_5 + s_7; S_2(C, V) = s_3 + s_2 + s_5 + s_6 + s_7$ Hence, we can deduce that: $S(C, U, V) = \frac{S_2(U,C)+S_2(C,V)-S_1(U,V)}{2}$.

We have described the necessary geometric formulas which are utilized in the paper. In the next sections, we present our approach to retrieve a human motion data based on the curves matching.

4 Human Motion Data Mining Based on the Curves Matching Approach

As we describe above, there are three main aspects concerned to the retrieval of the human motion data included the representation of the human motion data, the chosen distance function to measure the similarity and the considered techniques to speed up the searching. In Section 4 and Section 5, we investigate to describe our approach for these three aspects.

4.1 Spatial-Temporal Invariance Representation for Human Motion Capture Data

A simplified kinematic chain model is defined as Figure 5, which contains 14 joints that are constructed as a tree diagram. Eleven bones are extracted as the objects to represent motion feature, including 10 bones in the human limbs and a central bone connected by *root* and *chest* joints as a reference bone. Each bone is defined as a vector from the upper joint to the lower joint in the human skeleton. For every limb bone, the bone angle is defined as the angle between the limb bone and the central bone.

$$\theta_i^k = \cos^{-1}\left(\frac{\boldsymbol{B}_i^{(k)}.\boldsymbol{B}_i^{(Center)}}{\left|\boldsymbol{B}_i^{(k)}\right| \times \left|\boldsymbol{B}_i^{(Center)}\right|}\right), k = 1..10 \qquad (5)$$

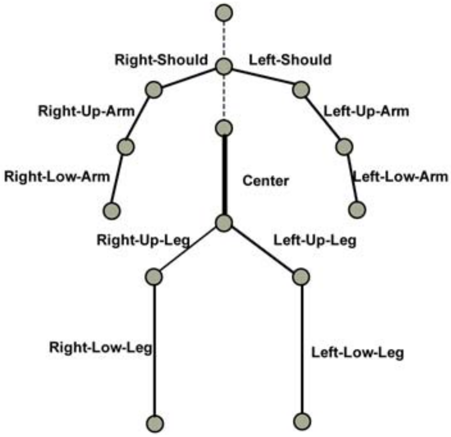


Fig. 5. Human skeleton and segments separated

where $\boldsymbol{B}_i^{(Center)}$ represents the central bone at the ith frame and θ is in the interval $[0, \pi]$. Consequently, by using the bone angle, we obtain a spatial invariance representation for human motion data. In addition, from Formula 5 the motion hereafter could be treat as a cluster of ten curves, in which each curve describes a bone motion. Hence, we can introduce the human motion M data as follows:

$$M = \left\{C^k(\theta_1^k, \theta_2^k, ...\theta_{m_k}^k)\right\}, k = 1..10 \tag{6}$$

In fact, the curves C_k are processed independently to carry out the technique of lower bounding or the technique of the dimensional simplification. Therefore, the curves C_k are considered as the temporal invariance representation of the human motion data. With such representation, our approach is a spatial-temporal invariance method. In practice, instead of using the motion M hereafter we utilize the curve C to describe our approach more easily. More, the similarity of the human motions could be treated as the similarity of the curves or the curves matching problem.

4.2 Distance Function to Measure the Similarity between Two Human Motions: Constrained Dynamic Time Warping Technique

Our chosen distance function to measure the similarity of the human motion is based on the constraint Dynamic Time Warping ($cDTW$) technique which is introduced in many time-series data matching references [14][16]. The $cDTW$ technique can be described as follows. Given two time-series curves due to the formula 1 $U(m) = (u_1, u_2, ..., u_m)$and $V(n) = (v_1, v_2, ..., v_n)$ and the time warping constraint ε, the constrained Dynamic Time Warping $cDTW$ is defined recursively as follows [16]:

$$
\begin{aligned}
&Dist_r(u_i, v_j) &&= \begin{cases} D_{base}(u_i, v_j)\,\text{if}\,|i-j| \leq r \\ \quad\infty \qquad\quad \text{otherwise} \end{cases} \\
&cDTW(\phi, \phi, r) &&= 0 \\
&cDTW(U, \phi, r) &&= cDTW(\phi, V, r) = \infty \\
&cDTW(U, V, r) &&= Dist_r(First(U), First(V)) + \\
& && \quad + \min \begin{cases} cDTW(U, Rest(V), r) \\ cDTW(Rest(U), V, r) \\ cDTW(Rest(U), Rest(V), r) \end{cases}
\end{aligned} \tag{7}
$$

where $r = \varepsilon \times t_{u,m} = \varepsilon \times t_{v,n}$ and ϕ is the empty curve, $First(U) = u_1$, $Rest(U) = u_2, u_3, ...u_m$ and D_{base} denotes the distance between two vertices of the curve. In our work, we utilize the Manhattan distance which is a special case of the Lm distance with index value α is equal to 1 defined as follows:

$$
\mathrm{L}_{\alpha=1}\ (\mathrm{u_i}(\mathrm{p_{u,i}}, \mathrm{t_{u,i}}),\ \mathrm{v_j}(\mathrm{p_{v,j}}, \mathrm{t_{v,j}})) = \Big(\ |\mathrm{p_{u,i}} - \mathrm{p_{v,j}}|^{\alpha=1} + |\mathrm{t_{u,i}} - \mathrm{t_{v,j}}|^{\alpha=1}\Big)^{\frac{1}{\alpha}=1} \tag{8}
$$

4.3 Lower Bounding Technique for Speeding Up the Human Motion Data Retrieval

As we describe previously, the main purpose of the using the lower bounding technique is to filter more quickly the negative time-series curves. Our lower bounding function is based on the LB_Keogh described as follows. Given a time-series curve as the formula 1, $C(m) = \{c_i(p_{c,i}, t_{c,i})\}, i = 1..m; t_i > t_j \,\forall i > j$ and a local constraint time warpingε, we note that $r = \varepsilon \times t_{c,m}$ and we use the term r to define two new time-series curves, U(m) and L(m):

$$
\begin{aligned}
&U(m) = \{u_i(p_{u,i}, t_{u,i})\}\ and\ L(m) = \{l_i(p_{l,i}, t_{l,i})\} \\
&where\, t_{u,i} = t_{l,i} = t_{c,i} \\
&p_{u,i} = \max(p_{c,i-r} : p_{c,i+r}) \\
&p_{l,i} = \min(p_{c,i-r} : p_{c,i+r})
\end{aligned} \tag{9}
$$

U(m) and L(m) stand for Upper and Lower, respectively. Figure 6 illustrates the created U and L with ε=20% of the length of the curve C. As we can see in the figure, U and L form a bounding envelope curves to enclose C from above to below. An important property of the bounding curves U and L is the following:

$$
\forall i, p_{u,i} > p_{c,i} > p_{l,i} \tag{10}
$$

Having defined the bounding curves U and L of the curve C and the Manhattan distance defined in Formula 8, we now use them to define a lower bounding measure for cDTW. Given two time-series curves C(m) and V(n), the bounding measure of V and the bounding envelopes of C are defined as follows:

$$
LB\_Keogh(C, V) = \sum_{i=1}^{t_{u,m}} \begin{cases} |p_{v,i} - p_{u,i}|\ if\, p_{v,i} > p_{u,i} \\ |p_{v,i} - p_{l,i}|\ if\, p_{v,i} < p_{l,i} \\ \quad 0 \quad otherwise \end{cases} \tag{11}
$$

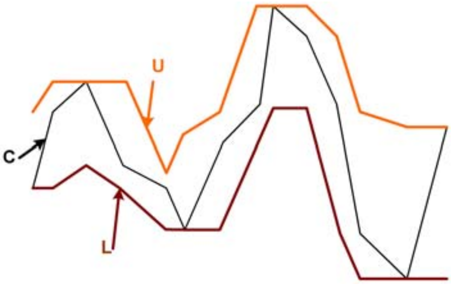


Fig. 6. An illustration of the time-series curves U and L, created for the time-series curve C(m) by using the LB_Keogh lower bounding function in which, r is a constant value due the length of $C(m) : r = 0.2 \times m$

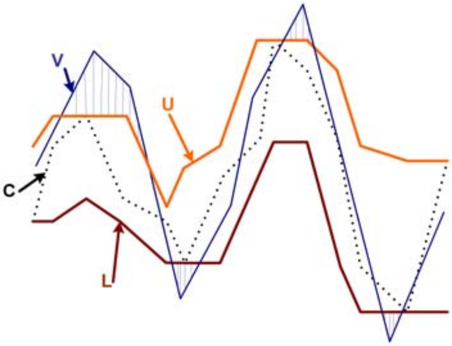


Fig. 7. An illustration of the lower bounding function LB_Keogh between the original curve C (shown dotted) and the query curve V calculated by the area of the outlier parts generated by the curve V and the none-self-intersecting polygon constituted by the bounding envelope U and L

This function can be visualized as the Manhattan distance between any parts of the query matching curve not falling within the envelope and the nearest (orthogonal) corresponding section of the envelope. Thus, the function is the area of the outlier parts generated by the curve V and the none-self-intersecting polygon constituted by two curves U and L, illustrated in Figure 7. Hence, we recall Formula 4 to calculate the LB_Keogh(C,V) in practice:

$$LB\_Keogh(C,V) = S(V,U,L) = \frac{S_2(U,V) + S_2(V,L) - S_1(U,L)}{2} \tag{12}$$

Now we will explain how utilize the LB_Keogh function to filter the negative curves to speed up the searching. In prior, we recall the work of Keogh et al. [16] which proved that:

$$LB\_Keogh(C,V) \leq cDTW(C,V) \tag{13}$$

Given a curve C and a threshold δ, we define a curves V as the positive matching curve with C if the inequality as follows is convenient:

$$cDTW(C, V) \leq \delta \tag{14}$$

Hence, we can see that the curve V will be the negative curve if

$$LB\_Keogh(C, V) > \delta. \tag{15}$$

In general, that inequality is utilized to filter the negative curves but our work will be continued for speed up the human motion data mining.

In this section, we have described an overall view of the human motion data mining in which, we conducted the human motion data mining as the curves matching problem by using the spatial-temporal representation for the human motion data. We also have presented the cDTW to measure the similarity of the curves and the technique to speed up the searching based on the lower bounding function called LB_Keogh for filtering rapidly the negative curves. In the next section, we describe another technique to speed up the searching based on the dimensional reduction technique called the curve simplification.

5 Techniques of Dimensionality Reduction Based on the Curve Simplification

As we describe above, the technique to reduce the dimension of the curves is utilized to speed up the searching besides the technique of using the lower bounding. In the approach of the curves matching problem, we utilize the curve simplification technique to reduce the dimension. The technique is proposed by Lowe [26] and is utilized efficiently to the key-frame extraction of the human motion data [1][5][27]. Due to the fact that, in our work, two axis of the coordinate system for the time-series curves do not have the same unit, while vertical axis is represented in the degree unit the horizontal one is represented in the frame number unit. Therefore instead of using the Euclidean distance of the point to a line as in the original version of Lowe technique, we utilize a distance measure of the length of the segment defined by the given point and the point lied in the line which has the same frame number value (orthogonal point). The idea is illustrated in Figure 8.

5.1 Lowe Algorithm to Simplify the Candidate Curve

Step 1: *Set the first and the last points as key points, creating 2 key points.*
Step 2: *Find the highest absolute distance point, which would have the distance called the error distance. If the error distance is lower than a specified threshold*σ*, stop further subdividing this interval. Otherwise, create a new key point at the point with the highest error distance.*

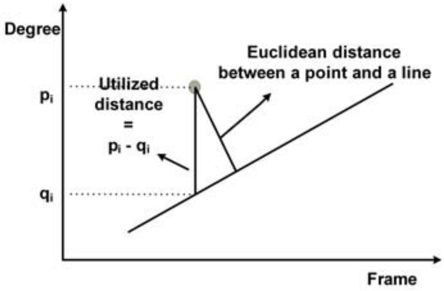


Fig. 8. An illustration of the utilized distance instead of using the Euclidean distance as in the original version of the Lowe technique

Step 3: *Sub-divide the current state into two smaller segments, a segment between the beginning key point and the newly created middle key point, and another segment between the middle key point and the ending key point. Assume that the new segments now have only 2 key points and restart the algorithm from step 2 for both segments.*

Figure 9 shows the modified algorithm to simplify the curve step by step. In practice, that modified algorithm is applied to the original candidate curves. Given a time-series curve C(m) with m is the dimension of the curve, we note $\bar{C}(\bar{m})$ is the simplified curve of C with the dimension of $\bar{m}$. In general, we always maintain the important property of the any time-series curve C(m) which is the tc,1 or tc,m referring to the same value.

In the next part, we will describe another modified Lowe algorithm to simplify the curve which adapts the bounding property of the envelope curves.

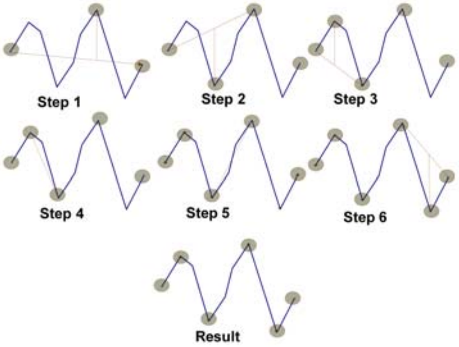


Fig. 9. An illustration of the modified curve simplification in which, the chosen points are highlighted

5.2 Modified Lowe Algorithm to Simplify the Bounding Envelope Curves

Opposed to the modified algorithm to simplify the When we simplify the bounding curves we have to ensure that the original curve is always enclosed by the bounding curves. Therefore, we proposed another modified Lowe algorithm for bounding curves simplification. We recall the utilized distance illustrated in Figure 10 in which, we occupy the value with the sign of the subtraction $p_i - q_j$. We differentiate into two cases, one for the upper bounding curve U and another for the lower bounding curve L.

Modified simplification technique for upper bounding curve is described as follows:

Step 1: *Set the first and the last points as key point, creating 2 key points.*

Step 2: *Find the highest positive distance point. If the point is found, create a new key point at the point with the highest positive distance; otherwise find the lowest negative distance point which would have the distance called the error distance. If the absolute value of the error distance is lower than a specified thresholdσ, stop further subdividing this interval. Otherwise, create a new key point at the point with the lowest negative error distance.*

Step 3: *Sub-divide the current state into two smaller segments, a segment between the beginning key point and the newly created middle key point, and another segment between the middle key point and the ending key point. Assume that the new segments now have only 2 key points and restart the algorithm from step 2 for both segments.*

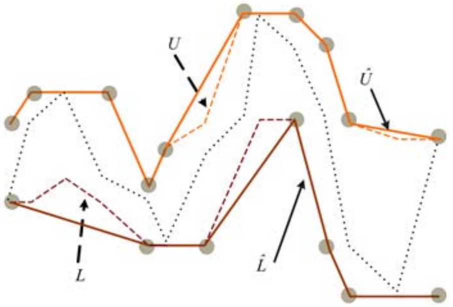


Fig. 10. Illustrates an example of the carrying out the modified curve simplification technique to the bounding curves, in which we note the simplified curve of the upper bounding as $\hat{U}$ and that of the lower bounding as $\hat{L}$

Modified simplification technique for lower bounding curve is described as follows:

Step 1: *Set the first and the last points as key point, creating 2 key points.*

Step 2: *Find the lowest negative distance point. If the point is found, create a new key point at the point with the lowest negative distance; otherwise find* *the highest positive distance point which would have the distance called the error distance. If the error distance is lower than a specified thresholdσ, stop further subdividing this interval. Otherwise, create a new key point at the point with the highest positive error distance.*

Step 3: *Sub-divide the current state into two smaller segments, a segment between the beginning key point and the newly created middle key point, and another segment between the middle key point and the ending key point. Assume that the new segments now have only 2 key points and restart the algorithm from step 2 for both segments.*

5.3 Pruning Power of the Human Motion Capture Data Mining

Having defined the dimensionality reduction of the candidate curves and the bounding envelope curves of the query curves, now we explain how to utilize them for pruning power of the human motion capture data mining. We recall Formula 12 to calculate the LB_Keogh of the query curve C and the candidate curve V in which, $LB_Keogh(C, V) = S(V, U, L)$is the area of the outlier parts generated by the curve V and the non-self-intersecting polygon constituted by two bounding curves of C. We occupy the simplified curve $\bar{V}$, the simplified curves $\hat{U}$ and $\hat{L}$to formulate a new bounding measure denoted LB_Keogh_Lowe(C,V) which is describe by the formula:

$$LB\_Keogh\_Lowe(C, V) = S(\bar{V}, \hat{U}, \hat{L}) = \frac{S_2(\hat{U}, \bar{V}) + S_2(\bar{V}, \hat{L}) - S_1(\hat{U}, \hat{L})}{2} \quad (16)$$

The difference between LB_Keogh and LB_Keogh_Lowe is illustrated in Figure 11. In practice, the computation cost of LB_Keogh_Lowe which performs the curve simplification on both of the bounding curves and the query curve is lower than that of the LB_Keogh.

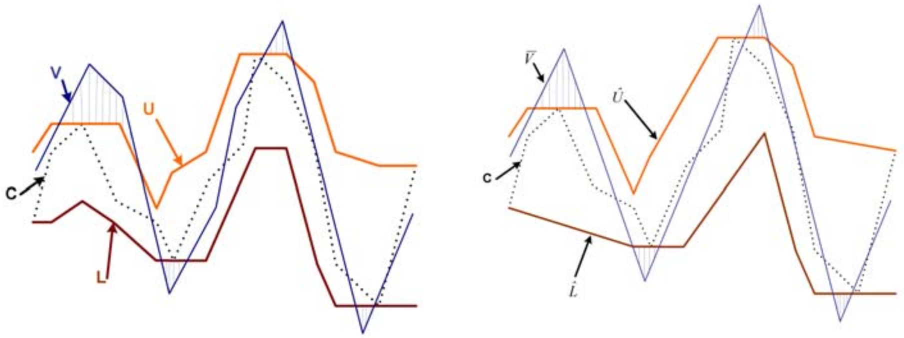
To filter rapidly the negative curves, given a threshold δ, instead of using the inequality 15 we utilize another inequality due to the LB_Keogh_Lowe in which, the query curve V will be a negative curve if

$$LB\_Keogh\_Lowe(C, V) > \delta \quad (17)$$

In practice, given a dataset of the human motions, we perform a pre-processing to obtain the dataset of query curves C, the lower bounding curves U and V, the reduced dimensionality lower bounding curves $\hat{U}$and $\hat{V}$. That pre-processing is considered as the offline working. Hence, it does not impact the performance of the system. More, we prefer using the δ as the function of the $t_{c,m}$ which means $\delta := \delta \times t_{c,m}$ due to the convenience of choosing δ.

We summarize this section by Figure 12 illustrating the proposed algorithm to handle the human motion data best match retrieval based on the curves matching.

In next section, we describe our experimental evaluation of our proposed approach with various real human motion datasets.



LB_Keogh lower bounding distance

LB_Keogh_Lowe lower bounding distance

Fig. 11. Comparison of LB_Keogh lower bounding distance (upper visual diagram): $LB\_Keogh(C,V) = S(V,U,L) = \frac{S_2(U,V)+S_2(V,L)-S_1(U,L)}{2}$ and LB_Keogh_Lowe lower bounding distance (lower visual diagram): $LB\_Keogh\_Lowe(C,V) = S(\bar{V},\hat{U},\hat{L}) = \frac{S_2(\hat{U},\bar{V})+S_2(\bar{V},\hat{L})-S_1(\hat{U},\hat{L})}{2}$

6 Experiments and Results

This section describes the experiments carried out to analyze the effectiveness of the curve simplification technique in term of the dimensionality reduction and the proposed lower bounding function which is based on that technique. In addition, we perform an experiment to verify the efficiency of the human motion data mining. The experiments were executed on an Intel Pentium Dual-Core 2.1GHz Window PC with 2GB RAM and the program is written in C++ Language.

To analyze the performance of the dimensionality reduction technique based on the curve simplification, we define a pruning scale P_1 which is described as follows:

$$P_1 = \frac{Number\,of\,the\,points\,of\,the\,reduced\,curves}{Number\,of\,the\,point\;\,of\,the\,original\,curves\,in\,database} \times 100(\%) \qquad (18)$$

To analyze the effectiveness of the proposed lower bounding function based on the curve simplification technique, we define another pruning power scale P_2 described as follows:

$$P_2 = \frac{Number\,of\,the\,human\,motions\,that\,do\,not\,require\,full\,wDTW}{Number\,of\,the\,human\,motions\,in\,databased} \times 100(\%) \qquad (19)$$

6.1 Data Preparing

We gathered more than 500Mb of the human motion dataset from Mocap-Club.com [31] with various kinds of human motion activities such as the walking, running, dancing, kung-fu and etc which are recorded in the frequency of 60Hz

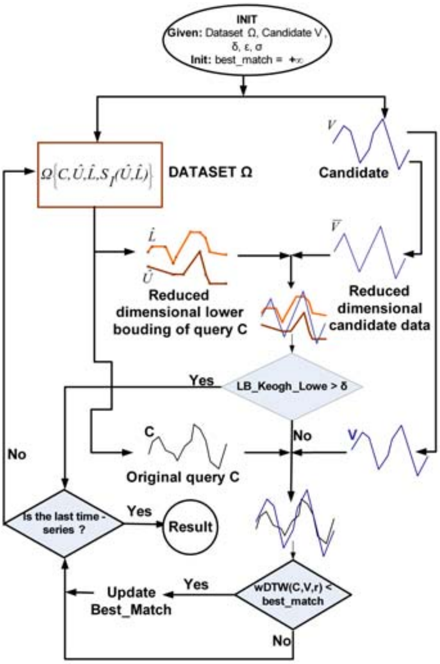

INIT
Given: Dataset Ω, Candidate V, δ, ε, σ
Init: best_match = +∞
$\Omega\{C,\hat{U},\hat{L},S_I(\hat{U},\hat{L})\}$
DATASET Ω
V
Candidate
$\hat{L}$
$\hat{U}$
Reduced dimensional lower bouding of query C
$\bar{V}$
Reduced dimensional candidate data
Yes
LB_Keogh_Lowe > δ
No
C
Original query C
No
V
Is the last time - series ?
Yes
Result
Yes
Update Best_Match
wDTW(C,V,r) < best_match
No

Fig. 12. An illustration of the algorithm to best match retrieval of the human motion data based on the curves matching

with more than 42 markers attached on the human body.Due to the proposed approach of the human motion representation, we extracted from the original dataset the 3D position of 14 joints of the kinematic chain model as described in Figure 5. Thus, we transformed them to the bone angle space and sort them so that each human motion is described by 10 curves of the bone angle. The obtained dataset differed significantly in size and length. In order to produce meaningful results, we reorganized obtained dataset in the bone space. We derived six sets of data, each containing 270 human motion sequences, with variable lengths of 32, 64, 128, 256, 512 and 1024, respectively. Short sequences were gathered by using a scale factor of the length of original data and the desired lengths of extracted data while long sequences were produced by concatenating original sequences. In practice, all experiments were conducted on these derived dataset.

6.2 Experiments and Results

Performance of the Dimensionality Reduction Technique Based on the Curve Simplification. Figure 13(a, b, c) illustrate the performance of the curve simplification P1 due to the length of the human motion. The Figure 13a is the outcome when we carry out the technique for the original dataset while Figure 13b and Figure 13c are the results carried out for the lower bounding curves. As

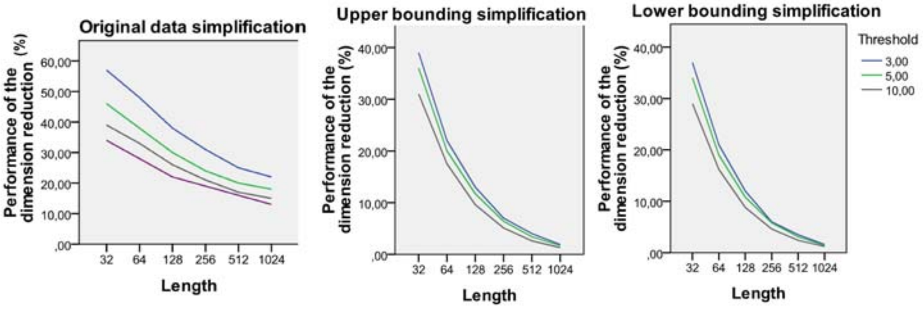


Fig. 13. (a, b, c) An illustration of the performance of the dimensionality reduction technique to the real human motion dataset. Fig 13a is outcome carried out the original human motion; Fig 13b, 13c are results carried out the upper bounding and lower bounding curves, respectively. The specified threshold is in degree.

shown in the figure, the performance of the curves simplification technique increased with the length of the data suggesting. In the lower bounding curves, the differences performances among the specified thresholds are not considerable thus we should choose the specified threshold as small as possible. In practice, to analyze the pruning power performance and the query time, we choose the threshold for lower bounding curves as $\sigma = 5^0$. The specified threshold to reduce the dimension of the original candidate data is chosen due to the threshold δ of the LB_Keogh_Lowe. We recall the work of Baek et al. [32] which developed experiment to choose the threshold for the human joint of the kinematic chain model. From this work, the suggested error interval for human joint is (100,200) thus we choose the specified threshold $\sigma = 10^0$ which is illustrated as the green curve in Figure 13a. Moreover, we also choose the threshold $\delta$of the LB_Keogh_Lowe 100 due to the reference work of Baek.

Pruning Power Performance. Having defined the threshold σ and δ, now we can verify the pruning power performance of the LB_Keogh_Lowe bounding function due to the capability of filtering the negative human motions. In fact, we perform the 1-nearest neighbor search using the sequential scan technique. A random human motion was chosen from the data set to act as the query and the remaining 269 human motions acted as the data. The search carried out 50 trials on each different lengths of the human motion. We recall Formula 20 to measure the pruning power for query human motion. The average of the 50 queries was reported as the pruning power of each different lengths of the dataset.

Figure 14 shows how the pruning power averaged of the proposed LB_Keogh_Lowe. Lower bounding function varies as the lengths of the data in the human motion dataset in which, 92% of the human motion of length 1024 and 67% of the human motion of length 32 did not require computation of the actual time warping distances. The promising pruning power greatly reduces the querying time. We conducted experiments to measure the time required for the query

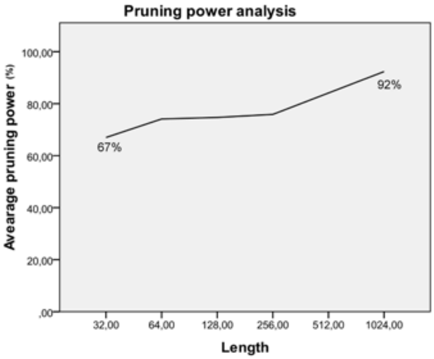


Fig. 14. Average pruning power in the human motion dataset

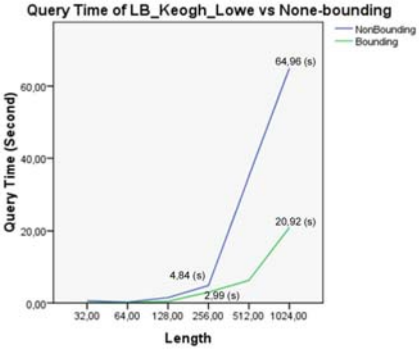


Fig. 15. Comparison the average query time between LB_Keogh_Lowe bounding and none bounding technique

evaluation of the human motion in different lengths. In prior, we should recall the work of Keogh et al. [16] to confirm the time warping constraintε. Due to the experimental result of Keogh the threshold $\varepsilon$is the most efficient with the value of 20% of the length of the real time-series data. Thus we denote ε=0.2.

Query Time Analysis. Having defined of $\sigma, \delta, \varepsilon$, we carried out the experiment to analyze the query time of the human motion. Similar to the pruning power analysis, the average value of 50 trials of each different lengths of the human motion are reported. Figure 15 illustrates the compared results of query time analysis in the human motion retrieval between the lower bounding approach and none-lower bounding approach. In fact, the average query time is consistently reduced and the difference between two query time curves is considerable, in which to search the 1-nearest, the human motion of length 256, our approach just needs 2.99 seconds averaged while the none-lower bounding approach needs 4,84 seconds averaged, thus we achieved 38% more efficient in time cost processing. To search the longest human motion of length 1024, our approach needs 20,92 second averaged while the none-lower bounding approach needs 64,96 seconds averaged. Hence, we also achieved 68% more efficient in time cost processing.

7 Conclusion and Future Work

In our paper, we have proposed a novel approach to retrieve a human motion data. The approach is based on the computational geometry and conducted the human motion data mining to the curves matching problem. We have introduced a spatial-temporal invariance representation for the human motion data. We have described the modified curve simplification algorithms adapted to the purpose of the human motion data mining and proposed the use of a lower bounding technique noted LB_Keogh_Lowe to speed up the searching. We have demonstrated the efficiency and effectiveness of our method on various real human motion datasets. Due to the obtained results, our approach can be applied in others contexts of the time-series data retrieval. For future work, we would like to carry out our approach to various kinds of time-series data to confirm the efficiency and the effectiveness and to compare with the other approaches in terms of the time-series data approximating representations and lower bounding distances.

Acknowledgments

The data used in this paper is obtained from http://www.mocapclub.com http://www.mocapclub.com.This work is part of SimAction project supported by the French Agency for the Research (ANR) through the call "Technologies for Health".

References

1. Muller, M., Roder, T., Clausen, M.: Efficient Indexing And Retrieval of Motion Capture Data Based on Adaptive Segmentation. In: Proceedings of the 4th Intl. Workshop on Content-Based Multimedia Indexing, Riga, Latvia (2005)
2. Giese, M.A., Poggio, T.: Morphable models for the analysis and synthesis of complex motion pattern. International Journal of Computer Vision 38(1), 59–73 (2000)
3. Kovar, L., Gleicher, M.: Automated extraction and parameterization of motions in large data sets. In: Proceedings of ACM SIGGRAPH (2004)
4. Keogh, E., Palpanas, T., Zordan, V.B., Gunopulos, D., Cardle, M.: Indexing large human-motion databases. In: Proceedings of the 30th VLDB Conference, Toronto, Canada, pp. 780–791 (2004)
5. Xiao, J., Zhuang, Y., Yang, T., Wu, F.: An efficient keyframe extraction from motion capture data. In: Advances in Computer Graphics, pp. 494–501 (2006)
6. Lee, J., Chai, J., Paul, S., Reitsma, A., Jessica, K.: Interactive Control of Avatars Animated with Human Motion Data. In: Proceedings SIGGRAPH, San Antonio, Texas, pp. 491–500 (2002)

7. Chui, C.Y., Chao, S.P., Wu, M.Y., Yang, S.N., Lin, C.: Content-based Retrieval for Human Motion Data. Journal of Visual Communication and Image Representation 16(3), 446–476 (2003)
8. Croitoru, A., Agouris, P., Stefanidis, A.: 3D trajectory matching by pose normalization. In: Geographic Information Systems Proceedings of the 13th Annual ACM International Workshop on Geographic Information Systems Bremen, Germany (2005)
9. Porikli, F.: Trajectory Distance Metric Using Hidden Markov Model based Representation. In: Proceeding of IEEE European Conference on Computer Vision (2004)
10. Berndt, J., Clifford, J.: Using dynamic time warping to find patterns in time series. In KDD Workshop (1994).
11. Chen, L., Ng, R. T.: On the marriage of lp-norms and edit distance. In VLDB (2004)
12. Chen, L., Ozsu, M. T., Oria, V.: Robust and fast similarity search for moving object trajectories. In: SIGMOD Conference (2005)
13. Chen, Y., Nascimento, M.A., Ooi, B.C., Tung, A.K.H.: SpADe: On Shape-based Pattern Detection in Streaming Time Series. In: ICDE (2007)
14. Keogh, E.: Exact Indexing of Dynamic Time Warping. In: Proceeding of VLDB Conference (2007)
15. Ding, Trajcevski, H., Scheuermann, G., Wang, P., Keogh, E.: Querying and Mining of Time Series Data Experimental Comparison of Representations and Distance Measures. In Proc. of the 34th VLDB, pp. 1542–1552 (2008)
16. Keogh, E., Ratanamahatana, C.A.: Exact indexing of dynamic time warping. Knowl. Inf. Syst.,7(3) (2005)
17. Keogh, E., Kasetty, S.: On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration. Data Min. Knowl. Discov., 7(4) (2003)
18. Faloutsos, C., Ranganathan, M., Manolopoulos, Y.: Fast Subsequence Matching in Time-Series databases. In: SIGMOD Conference (1994)
19. Chan, K.P., Fu, A.W.-C.: Efficient Time Series Matching by Wavelets. In: ICDE (1999)
20. Keogh, E., Chakrabarti, K., Pazzani, M.J., Mehrotra, S.: Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. Knowl. Inf. Syst., 3(3) (2001)
21. Keogh, E., Chakrabarti, K., Mehrotra, S., Pazzani, M.J.: Locally Adaptive Dimensionality Reduction for Indexing Large Time Series Databases. In: SIGMOD Conference (2001)
22. Keogh, E., Chakrabarti, K., Pazzani, M.J., Mehrotra, S.: Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. Knowl. Inf. Syst., 3(3) (2001)
23. Cai, Y., Ng, R.T.: Indexing spatio-temporal trajectories with chebyshev polynomials. In: SIGMOD Conference (2004)
24. Lin, J., Keogh, E., Wei, L., Lonardi, S.: Experiencing SAX: a novel symbolic representation of time series. Data Min. Knowl. Discov., 15(2) (2007)
25. Chen, Q., Chen, L., Lian, X., Liu, Y., Yu, J.X.: Indexable PLA for Efficient Similarity Search. In: VLDB (2007)
26. Lowe, G.: Three-dimensional object recognition from single two dimensional images. Artificial Intelligence 31(3), 355–395 (1987)
27. Lim, L.S., Thalmann, D.: Key-posture extraction out of human motion data by curve simplification. In: Proceedings of the 23rd Annual EMBS International Conference, Istanbul, Turkey (2001)

28. Onder, O., Gudukbay, U., Ozguc, B., Erdem, T., Ozkan, C.: Keyframe Reduction Techniques for Motion Capture Data. In: Proceeding of 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (2008)
29. Kim, S., Park, S., Chu, W.: An Index-based approach for similarity search supporting time warping in large sequence databases. In: Proc. 17th International Conference on Data Engineering, pp. 607–614 (2001)
30. Yi, B., Jagadish, K., Faloutsos, H.: Efficient retrieval of similar time sequences under time warping. In: ICDE, pp. 23–27 (1998)
31. http://www.mocapclub.com/Pages/Library.htm
32. Baek, S., Lee, S., Kim, G.J.: Motion retargeting and evaluation for VR-based training of free motions. Journal of Visual Computer 19(4), 222–224 (2003)
33. http://softsurfer.com/Archive/algorithm_0101/algorithm_0101.htm

Estimation System for Forces and Torques in a Biped Motion

Andrzej Polański[1,2], Adam Świtoński[1,2], Henryk Josiński[1,2],
Karol Jędrasiak[1,2], and Konrad Wojciechowski[1,2]

[1] Polish Japanese Institute for Information Technology, Bytom, Poland
andrzej.polanski@polsl.pl
http://bytom.pjwstk.edu.pl
[2] Institute of Informatics Silesian Technical University
Akademicka 16, 44-100 Gliwice, Poland
http://bytom.pjwstk.edu.pl

Abstract. In this paper we present a simple estimation system for ground reaction forces, inter-segmental forces and joint torques in the biped motion. The proposed system bases on classification of walk into two phases, left leg stance - right leg swing, and left leg swing - right leg stance. The system does not need any additional measurements of forces and moments. It uses matrix formulation for inverse dynamics problem of computing moment and torques on the basis of measurements of velocities and accelerations. The input signals are filtered with the use of the Savitzky Golay smoothing filter.

1 Introduction

Studies on human motion, involving e.g.,estimation of masses and moments of inertia of body parts, modeling dynamics of body parts and their joints [3], [6], estimating forces and torques in human body motions [5], simulations of human motions [17], correlating measurements of muscle tensions and parameters of motion, and others, are of great importance in many areas of science related to biomechanics and neurophysiology [11], [13],[19].

This paper is devoted to the problem of estimation of ground reaction forces, inter-segmental forces and joint torques by using the method of inverse dynamics. This problem has been very extensively studied by many authors, [1], [2], [4], [5], [7], [9], [8], [10], [16], [15]. The idea of the inverse dynamics approach is to first estimate positions, angles, velocities and accelerations of elements of a mechanical system, and then use these estimates to formulate systems of algebraic equations to be solved for forces and moments. There are, however, several formulations of the inverse dynamics method, which differ in assumptions made in constructions of inverse dynamics equations and in the methodologies used for solving these systems of equations. Concerning constructions of inverse dynamics equations, their research led to the observation of sensitivity of estimated forces and torques to errors in computed accelerations and velocities [9]. Due to this

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 185–192, 2010.

sensitivity, the quality of estimation of forces and torques is strongly influenced by the choice of the filtering algorithm, for reduction of the level of input noise [10]. Another point in constructions of inverse dynamics equations is studying the possibility of using some additional data, mainly measurements of ground reaction forces, to improve stability of solutions [18]. Concerning the methodologies used for solving inverse dynamics equations, two approaches are most often applied, top down or bottom up [8]. In both these approaches forces and moments are estimates on the basis of iterative analyses of elements of mechanical chains. In the top down method the iterations start from the free end of the mechanical chain, while in the bottom up method the iterations are initiated at the constrained end of the mechanical chain. A very important issue in systems for estimation inter-segmental forces and joint torques is verification of the results of computations versus experimental measurements. A most often applied verification paradigm uses quality indexes which integrate computed forces and torques with the measured ground reaction forces [5].

In this paper we describe an estimation system for the analysis of walk of a two dimensional biped, based on the inverse dynamics principle. Compared to several methodologies of applying inverse dynamics to estimation of forces and torques, mentioned above, the proposed system has three original features. First, the system allows for analysis of human motion data, which do not contain measurements of ground reaction forces. Since there are a lot of sources of such data available in the internet, our system is very useful for many types of analyses and comparisons. Second, in the proposed solution we do not use iterative (top down or bottom up) algorithms. Instead, a system of algebraic equations in the matrix form, including velocities and accelerations of all segments and all forces and torques is written, and estimates are obtained by solving this system, in the least square sense, for ground reaction forces, inter-segmental forces and joint torques. In this approach, errors are better balanced, compared to iterative bottom up or top down algorithms, where the propagation of error is not controlled. Also top down or bottom up approaches are not straightforward to apply for branched mechanical systems, as is in the case of a biped, analyzed here. The least square residual error is used as a measure of the quality of estimation. The third original feature is the use of the Savitzky-Golay (SG) filtration algorithm [12] to reduce input noise and to, therefore, improve estimates of velocities and accelerations. We demonstrate that this algorithm is well suited for the filtration step in algorithms for estimation of human motion, due to its flexibility and ability to cope with wide ranges of frequencies of noise.

We illustrate the proposed approach by using publicly available human walk data from the Carnegie Mellon University Motion Capture Database. The three dimensional data in the amc and asf formats were limited to five segment biped (consisting of two two - segment legs and a lower back) and turned into two dimensions by appropriate projections. The resulting motion data were then used in the inverse dynamics algorithm for estimation of ground reaction forces, inter-segmental forces and joint torques.

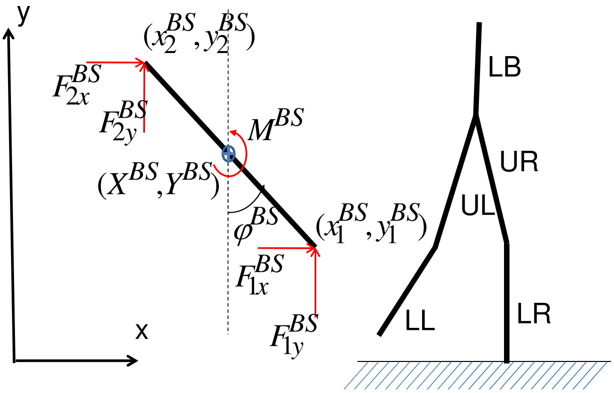


Fig. 1. Left - one body segment with notation for coordinates, forces and moments. Right - scheme of the biped

2 Mathematical Model of the Biped Motion

The model of the biped is presented in Fig. 1, in the right panel. It includes five body segments, with the following names and abbreviations (given in parentheses and shown in Fig. 1), lower back (LB), upper left leg (UL), lower left leg (LL), upper right leg (UR) and lower right leg (LR). The two-letter abbreviations are used as indexes in equations for dynamics of these segments of body. Additionally, BS is used as an abbreviation for "body segment", and stands as a "generic" index, which can mean any of the biped segments. In any specific context it can be replaced by one of the abbreviations listed above. In Fig. 1, in the left panel, one body (biped) segment is presented. The directions and orientations of coordinates x and y are also depicted. The following notation is used

X^{BS}, $\dot{X}^{BS}$, $\ddot{X}^{BS}$ - position, velocity and acceleration, along axis x, of the center of mass of the body segment,

Y^{BS}, $\dot{Y}^{BS}$, $\ddot{Y}^{BS}$ - position, velocity and acceleration, along axis y, of the center of mass of the body segment,

φ^{BS}, $\dot{\varphi}^{BS}$, $\ddot{\varphi}^{BS}$ - elevation angle, angular velocity and angular acceleration of the body segment,

x_1^{BS}, y_1^{BS}, x_2^{BS}, y_2^{BS} - x and y coordinates of the lower and upper ends of the body segment

m^{BS} - mass of the body segment,

I^{BS} - moment of inertia of the body segment,

l^{BS} - length of the body segment,

F_{1x}^{BS}, F_{1y}^{BS}, F_{2x}^{BS}, F_{2y}^{BS} - forces, along x and y axes acting on the body segment, respectively at the lower and upper end,

M^{BS} - moment acting on the body segment,

g - gravitational acceleration.

2.1 A Model for a Single Biped Segment

The 2D dynamics of a single biped (body) segment is described by the following system of three differential equations, equations of motion along coordinate x and along coordinate y

$$\ddot{X}^{BS} = \frac{F_{1x}^{BS} + F_{2x}^{BS}}{m^{BS}}, \ \ddot{Y}^{BS} = \frac{F_{1y}^{BS} + F_{2y}^{BS}}{m^{BS}} - g, \tag{1}$$

and equation for rotation

$$\ddot{\varphi}^{BS} = \frac{l^{BS}[(F_{1x}^{BS} - F_{2x}^{BS})\cos(\varphi^{BS}) + (F_{1y}^{BS} + F_{2y}^{BS})\sin(\varphi^{BS})]}{2I^{BS}} + \frac{M^{BS}}{I^{BS}}. \tag{2}$$

By using the following geometric relations between segment centers of gravity and segment ends:

$$x_1^{BS} = X^{BS} + \frac{l^{BS}\sin(\varphi^{BS})}{2}, \ y_1^{BS} = Y^{BS} - \frac{l^{BS}\cos(\varphi^{BS})}{2}, \tag{3}$$

$$x_2^{BS} = X^{BS} - \frac{l^{BS}\sin(\varphi^{BS})}{2}, \ y_2^{BS} = Y^{BS} + \frac{l^{BS}\cos(\varphi^{BS})}{2}, \tag{4}$$

and differentiating two times, we obtain equations for accelerations of segment ends, which we write in the vector form

$$\begin{bmatrix} \ddot{x}_1^{BS} \\ \ddot{y}_1^{BS} \\ \ddot{x}_2^{BS} \\ \ddot{y}_2^{BS} \end{bmatrix} = M_{coeff}(\varphi^{BS}, \dot{\varphi}^{BS}) \begin{bmatrix} F_{1x}^{BS} \\ F_{1y}^{BS} \\ F_{2x}^{BS} \\ F_{2y}^{BS} \\ M^{BS} \end{bmatrix} + \begin{bmatrix} -\frac{(\dot{\varphi}^{BS})^2 l^{BS}\sin(\varphi^{BS})}{2} \\ \frac{(\dot{\varphi}^{BS})^2 l^{BS}\cos(\varphi^{BS})}{2} - g \\ \frac{(\dot{\varphi}^{BS})^2 l^{BS}\sin(\varphi^{BS})}{2} \\ -\frac{(\dot{\varphi}^{BS})^2 l^{BS}\cos(\varphi^{BS})}{2} - g \end{bmatrix}. \tag{5}$$

The entries of the matrix of coefficients, $M_{coeff}(\varphi^{BS}, \dot{\varphi}^{BS})$ follow from computing second derivatives of (3)-(4).

2.2 Interactions Between Biped Segments and between the Biped and the Ground

In this subsection we aggregate equations of motions (5) for all biped segments to obtain the complete equations of the mathematical model for the biped motion. Let us introduce symbolic notation for vectors in (5) in order to shorten the expressions for the dynamics of the biped,

$$\ddot{\xi}^{BS} = \begin{bmatrix} \ddot{x}_1^{BS} \\ \ddot{y}_1^{BS} \\ \ddot{x}_2^{BS} \\ \ddot{y}_2^{BS} \end{bmatrix}, \ \Phi^{BS} = \begin{matrix} F_{1x}^{BS} \\ F_{1y}^{BS} \\ F_{2x}^{BS} \\ F_{2y}^{BS} \\ M^{BS} \end{matrix}, \ f^{BS} = \begin{bmatrix} -\frac{(\dot{\varphi}^{BS})^2 l^{BS}\sin(\varphi^{BS})}{2} \\ \frac{(\dot{\varphi}^{BS})^2 l^{BS}\cos(\varphi^{BS})}{2} - g \\ \frac{(\dot{\varphi}^{BS})^2 l^{BS}\sin(\varphi^{BS})}{2} \\ -\frac{(\dot{\varphi}^{BS})^2 l^{BS}\cos(\varphi^{BS})}{2} - g \end{bmatrix}. \tag{6}$$

Using the above vectors the whole model for the biped can be written in the block matrix form, shown below

$$\begin{bmatrix} \ddot{\xi}^{LB} \\ \ddot{\xi}^{UL} \\ \ddot{\xi}^{LL} \\ \ddot{\xi}^{UR} \\ \ddot{\xi}^{LR} \end{bmatrix} = \begin{bmatrix} M^{LB}_{coeff} & 0 & 0 & 0 & 0 \\ 0 & M^{UL}_{coeff} & 0 & 0 & 0 \\ 0 & 0 & M^{LL}_{coeff} & 0 & 0 \\ 0 & 0 & 0 & M^{UR}_{coeff} & 0 \\ 0 & 0 & 0 & 0 & M^{LR}_{coeff} \end{bmatrix} \begin{bmatrix} \Phi^{LB} \\ \Phi^{UL} \\ \Phi^{LL} \\ \Phi^{UR} \\ \Phi^{LR} \end{bmatrix} + \begin{bmatrix} f^{LB} \\ f^{UL} \\ f^{LL} \\ f^{UR} \\ f^{LR} \end{bmatrix} . \quad (7)$$

The above system of differential equations describes dynamics of all biped segments in a separated way. However, in reality biped segment are jointed together and form one dynamical system. All forces and moments, F^{BS}_{1x}, F^{BS}_{1y}, F^{BS}_{2x}, F^{BS}_{2y}, M^{BS}, which appear as element of vectors Φ^{BS} in equations (7) should be interpreted as ground reaction forces, inter-segmental forces or joint torques. We introduce the notation M^{LB_UL}, M^{LB_UR}, M^{UL_LL}, M^{UR_LR} for torques between joints, and $F_x^{LB\_UL}$, $F_y^{LB\_UL}$, $F_x^{LB\_UR}$, $F_y^{LB\_UR}$, $F_x^{UL\_LL}$, $F_y^{UL\_LL}$, $F_x^{UR\_LR}$, $F_y^{UR\_LR}$ for inter-segmental reaction forces along axes x and y between biped segments, and $F_x^{G\_LL}$, $F_y^{G\_LL}$, $F_x^{G\_LR}$, $F_y^{G\_LR}$ for ground reaction forces along axes x and y for the segment LL and LR.

When the dynamical system built of biped segments LB, UL, LL, UR, LR joined as shown in Fig. 1, is analyzed, the above torques and forces should replace forces and moments F^{BS}_{1x}, F^{BS}_{1y}, F^{BS}_{2x}, F^{BS}_{2y}, M^{BS} in equations (7). The rules for substituting reaction forces and torques in the system of equations (7) follow from the geometries in Fig. 1. For joint torques we have: $M^{LB} = M^{LB\_UL} + M^{LB\_UL}$, $M^{UL} = -M^{LB\_UL} + M^{UL\_LL}$, $M^{UR} = -M^{LB\_UR} + M^{UR\_LR}$, $M^{LL} = -M^{UL\_LL}$, $M^{LR} = -M^{UR\_LR}$.

For inter-segmental forces the replacements rules are as follows: $F^{LB}_{2x} = 0$, $F^{LB}_{2y} = 0$, $F^{LB}_{1x} = F_x^{LB\_UL} + F_x^{LB\_UR}$, $F^{LB}_{1y} = F_y^{LB\_UL} + F_y^{LB\_UR}$, $F^{UL}_{1x} = F_x^{UL\_LL}$, $F^{UL}_{1y} = F_y^{UL\_LL}$, $F^{UL}_{2x} = -F_x^{LB\_UL}$, $F^{UL}_{2y} = -F_y^{LB\_UL}$, $F^{UR}_{1x} = F_x^{UR\_LR}$, $F^{UR}_{1y} = F_y^{UR\_LR}$, $F^{UR}_{2x} = -F_x^{LB\_UR}$, $F^{UR}_{2y} = -F_y^{LB\_UR}$, $F^{LL}_{2x} = -F_x^{UL\_LL}$, $F^{LL}_{2y} = -F_y^{UL\_LL}$, $F^{LR}_{2x} = -F_x^{UR\_LR}$, $F^{LR}_{2y} = -F_y^{UL\_LR}$.

The last group of forces are ground reaction forces. These forces replace forces F^{LL}_{1x}, F^{LL}_{1y} and F^{LR}_{1x}, F^{LR}_{1y}. There are two variants of these replacement rules, depending on the phase of walk. The variant of the replacement rule depends on which leg is in the stance and which in the swing phase. For the phase LL-stance and LR-swing we have $F^{LL}_{1x} = F_x^{G\_LL}$, $F^{LL}_{1y} = F_y^{G\_LL}$, $F^{LR}_{1x} = 0$, $F^{LR}_{1y} = 0$, and for the phase LL-swing and LR-stance we have $F^{LL}_{1x} = 0$, $F^{LL}_{1y} = 0$, $F^{LR}_{1x} = F_x^{G\_LR}$, $F^{LR}_{1y} = F_y^{G\_LR}$.

By using the above replacement rules, we define a 20 × 14 binary connection matrix which relate all 20 body segments forces and moments to 14 ground reaction forces, inter-segmental forces or joint torques. For the walk phase LL-stance and LR-swing this relation has the form of the binary matrix, which we denote by $M^{LL-stanceLR-swing}_{replacements}$ and for the walk phase LL-swing and LR-stance the relation has the form of the binary matrix denoted by $M^{LL-swingLR-stance}_{replacements}$.

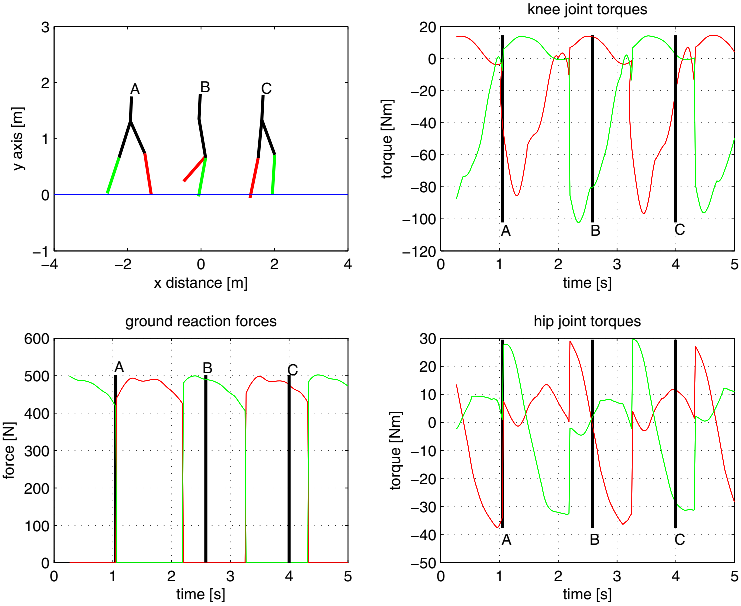


Fig. 2. Example of using the developed system for computing y coordinates of ground reaction forces, hip joint torques and knee joint torques

Entries of the matrices $M_{replacements}^{LL-stanceLR-swing}$ and $M_{replacements}^{LL-swingLR-stance}$ are defined by the replacement rules given above.

2.3 Least Squares Solutions for Forces and Torques

Substituting $M_{replacements}^{LL-stanceLR-swing}$ or $M_{replacements}^{LL-swingLR-stance}$ in the system of differential equations (7) we obtain equations which contain vector of unknown forces and moments versus estimated values of body segments elevation angles and their derivatives, and estimated accelerations of body segment ends. There are two separate equations, one for the walk phase LL-stance and LR-swing and another one for the walk phase LL-swing and LR-stance. By solving these systems of equations in the least squares sense we obtain estimates of ground reaction forces, inter-segmental forces and joint torques.

Estimates of signals and their derivatives are obtained by using the Savitzky Golay (SG) filter, a smoothing filter, which is very well suited for rejecting noise, when (i) the signal has rather wide range of the frequency spectrum or (ii) when there is a need to preserve as closely as possible positions and shapes of maxima or minima.

The additional advantage of the SG filters is that since the polynomial approximation is an element of the filtration algorithm, first, second (and if necessary higher) derivatives of the signal are also available as a result of filtration. Since application of the method of inverse dynamics requires computing second derivatives, we found SG filter a very useful tool.

3 Example

We have applied the above described methodology to the problem of estimation of ground reaction forces, inter-segmental forces and joint torques based on the publicly available human motion data from the Carnegie Mellon University Motion Capture Database (http://mocap.cs.cmu.edu/). We have downloaded and studied several films of walks, here we present results of analyses of the files '07.asf' and '07_01.amc'. We have limited the data to five body segments listed in the previous section, lower back, upper left leg, lower left leg, upper right leg and lower right leg. The data formats asf and amc in the Mocap database describe three dimensional motion. In order to project this data onto a 2D space, for each body segment its angle with axis y was computed and then x and y coordinates were found by following mechanical chain with the constraint that lengths of the body segments must not vary.

Samples of results of computations of ground reaction forces (y coordinates), hip joint torques and knee joint torques are presented in Fig. 2.

4 Conclusion

The presented estimation system for ground reaction forces, inter-segmental forces and joint torques for the motion of a biped, has the ability to estimate values of forces and torques, solely on the basis of the time sequence of the measured positions of body segments, without the need of using any additional measurements. There are many data sources of this type and the possibility for analyzing such data is useful in many researches. The system uses the principle of inverse dynamics, second derivatives of signals are obtained as output of the Savitzky - Golay filter, which additionally reduces the level of noise.

The system uses two models for the least squares formulation of the estimation problems, depending on the phase of walk, LL-stance - LR-swing or LL-swing - LR-stance. One can see that, especially when analyzing many motion data files, when using our approach, one encounters the problem of automatic classification of biped walk frames into walk phases LL-stance - LR-swing and LL-swing - LR-stance. This problem can be solved by using results of the research presented in the paper [14].

Acknowledgment

This paper has been supported by the project „System with a library of modules for advanced analysis and an interactive synthesis of human motion" co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme - Priority Axis 1. Research and development of modern technologies, measure 1.3.1 Development projects.

References

1. Anderson, F.C., Pandy, M.G.: Static and dynamic optimization solutions for gait are practically equivalent. Journal of Biomechanics 34, 153–161 (2001)
2. Cahou, V., Luc, M., David, A.: Static optimal estimation of joint accelerations for inverse dynamics problem solution. Journal of Biomechanics 35, 1507–1513 (2002)
3. Featherstone, R.: Rigid body dynamics algorithms. Springer, Heidelberg (2008)
4. Forner-Cordero, A., Koopman, H.J.F.M., van der Helm, F.C.T.: Inverse dynamics calculations during gait with restricted ground reaction force information from pressure insoles. Gait & Posture 23, 189–199 (2006)
5. Hatze, H.: The fundamental problem of myoskeletal inverse dynamics and its implications. Journal of Biomechanics 35, 109–115 (2002)
6. Ma, B., Wu, Q.: Parametric study of repeatable gait for a planar five-link biped. Robotica 20, 493–498 (2002)
7. Remy, C.D., Thelen, D.G.: Optimal Estimation of Dynamically Consistent Kinematics and Kinetics for Forward Dynamic Simulation. Journal of Biomechanical Engineering 131, 031005-1–031005-9 (2009)
8. Riemer, R., Hsiao-Wecksler, E.T.: Improving joint torque calculations: Optimization-based inverse dynamics to reduce the effect of motion errors. Journal of Biomechanics 41, 1503–1509 (2008)
9. Riemer, R., Hsiao-Wecksler, E.T., Zhang, X.: Uncertainties in inverse dynamics solutions: A comprehensive analysis and an application to gait. Gait & Posture 27, 578–588 (2008)
10. Runge, C., Zajac, F.E., Allum, J.H.J., Risher, D.W., Bryson, A.E., Honegger, F.: Estimating Net Joint Torques from Klnesiological Data Using Optimal Linear System Theory. IEEE Trans. Biomedical Engineering 42, 1158–1164 (1995)
11. Samson, W., Desroches, G., Cheze, L., Dumas, R.: 3D dynamics of healthy shildres's gait. Journal of Biomechanics 42, 2447–2453 (2009)
12. Savitzky, A., Golay, M.J.E.: Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry 36, 1627–1639 (1964)
13. Shelburne, K.B., Pandy, M.G., Anderson, F.C., Torry, M.R.: Pattern of anterior cruciate ligament force in normal walking. Journal of Biomechanics 37, 797–805 (2004)
14. Switonski, A., Josinski, H., Jedrasiak, K., Polanski, A., Wojciechowski, K.: Classification of poses and movement phases (submitted)
15. Thelen, D.G., Anderson, F.C., Delp, S.L.: Generating dynamic simulations of movement using computed muscle control. Journal of Biomechanics 36, 321–328 (2003)
16. Thelen, D.G., Anderson, F.C.: Using computed muscle control to generate forward dynamic simulations of human walking from experimental data. Journal of Biomechanics 39, 1107–1115 (2006)
17. Wang, J.M., Fleet, D.J., Hertzmann, A.: Gaussian Process Dynamical Models for Human Motion. IEEE Trans. PAMI 30, 283–298 (2008)
18. Zajac, F.E., Neptune, R.R., Kautz, S.A.: Biomechanics and muscle coordination of human walking, Part I: Introduction to concepts, power transfer, dynamics and simulations. Gait and Posture 16, 215–232 (2002)
19. Zajac, F.E., Neptune, R.R., Kautz, S.A.: Biomechanics and muscle coordination of human walking, Part II: Lessons from dynamical simulations and clinical implications. Gait and Posture 17, 1–17 (2003)

Classification of Poses and Movement Phases

Adam Świtoński[1,2], Henryk Josiński[1,2], Karol Jędrasiak[1],
Andrzej Polański[1,2], and Konrad Wojciechowski[1,2]

[1] Polish-Japanese Institute of Information Technology
Aleja Legionów 2, 41-902 Bytom, Poland
{aswitonski,apolanski,hjosinski,kwojciechowski}@pjwstk.edu.pl
[2] Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice Poland
{firstname.name}@polsl.pl

Abstract. We have focused on the problem of classification of motion frames representing different poses by supervised machine learning and dimensionality reduction techniques. We have extracted motion frames from global database manually, divided them into six different classes and applied classifiers to automatic pose type detection. We have used statistical Bayes, neural network, random forest and Kernel PCA classifiers with wide range of their parameters. We have tried classification on the original data frames and additional reduced their dimensionality by PCA and Kernel PCA methods. We have obtained satisfactory results rated in best case 100 percent of classifiers efficiency.

1 Introduction

Motion databases consist of a very large amount of data. They store hundreds of motions and each motion is a large sequence of frames, usually captured with minimum 100Hz frequency. Motion data usually comes directly from mocap capturing devices [8] or they can be estimated from static 2D images. In [6] such a method based on the Markov chain Monte Carlo is proposed. In practice it is impossible to analyze and search in such kind of databases manually. One of the first tasks in the automatic motion analysis could be pose detection, which means pointing for each frame of the pose type. It could be useful in database searching problem. On the basis of labeled frames we are able to build criteria of the database query, for instance find a motion in which a human is sitting. Pose identification also has medical applications. It could be used in automatic detection of some improper poses, typical in given kinds of diseases. Finally, pose identification could be useful in further automatic motion analysis, as for instance segmentation. The boundary of the motion segments could be placed in the moment of pose changing.

Comparing of motion frames is not trivial problem. A frame is described by position of special markers located by the joints. The frame can be represented by the data of direct position of each marker in 3D global coordinate system. But the most often used representation is a kinematic chain, which has format of a tree structure. The root object is placed on the top of the tree and is

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 193–200, 2010.

described by its position in global coordinate system. Child objects are connected to their parents and have information of transformation relative to the parents. Both formats contain exactly the same data, but the advantage of kinematic chain is that identical poses captured in different places have almost the same numerical representation, except for root objects, which is completely different in the representation in 3D global coordinate system [8].

On the basis of the frame format we can build pose similarity measures. The distance could be an aggregation of distances each pair of suitable markers in 3D global coordinate system, but it has disadvantages described above. Thus, in practice it is not used. The authors of the [4] propose 3D cloud point distance measure. First they build cloud points for compared frames and their temporal context. Further, they find global transition to match both clouds and finally calculate the sum of distances corresponding points of matched clouds. In [5] clouds are built based on the downsampled frame representation, which avoids focusing on the pose details.

In kinematic chain format, transformations are usually coded with unit quaternions. Thus pose distance can be evaluated as sum of quaternion distances. In [3] frame distance is total weighted sum of quaternion distances because influence of transformations can differ on the pose - the differences depends on the joints.

[7] propose binary relational motion features as description of pose. Relational feature is enabled if given joints and bones are in the defined relation: the left knee is behind the right knee, the right ankle is higher than the left knee and so on. We prepare such a set of features and this way describe pose by the binary vector. Pose distance can be calculated as distance metric of vectors descriptors. The basic problem in relational motion features is the proper set of features to distinguish between different kinds of poses. It is very difficult to prepare a single set of features which is applicable to the recognition of every kind of pose. Features are usually dedicated to specialized detections and because of their relatively easy interpretation they are prepared by medical experts who know the meaning of the given joint and bones dependencies. We can generate large features vectors from generic features set proposed by [7], but because of the difficulty in pointing significant features, this leads to long pose description and redundant data.

The problem of recognition of a pose type is much more general than evaluating similarity of two different poses. A single pose type can be represented by different, not similar poses. It is so because different phases of each pose type, for example jumping can be divided into starting, flying and landing phases. Secondly it happens so because of different characters of the same pose type - fast run and slow run generates other pose frames.

Each pose is represented by the location of tens of markers. Thus, a manual discovery of dependencies in each pose and finding numerical boundaries of given poses representation is almost impossible. Considering that we have decided to use machine learning techniques which are able to explore data, find dependencies and generalize knowledge. We have tested supervised learning with pose distance metric based on the tree like representation. We have also tested linear and nonlinear dimensionality reduction methods to reduce pose description.

2 Pose Database

We have prepared poses data from Carnegie Mellon University Motion Capture Database [2]. We have analyzed motion clips, selected pose frames of six different pose types: climbing, jumping, running, sitting, standing and walking. Each pose type contains a wide range of instances, taken from different movements and in different move phases. Finally we have 59761 labeled pose frames. Example poses are shown in Fig. 1.

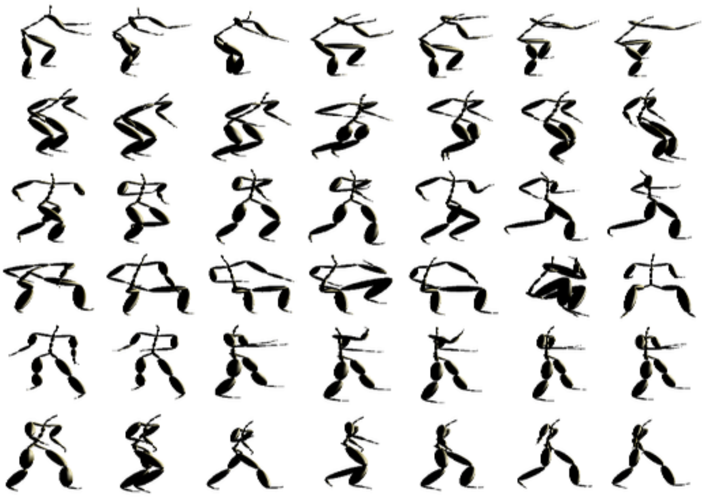


Fig. 1. Randomly selected poses from prepared test database. The following rows represent: climbing, jumping, running, sitting, standing and walking pose types.

Each pose is identified by six root attributes pointing location and orientation of global coordinate system and 56 relative attributes pointing twenty six body parts in a tree like structure. The number of description values of a given body part depends on its degrees of freedom.

In preprocessing step we have removed root attributes to avoid learning of classifiers pose type by location and orientation of global coordinate system. The data originates from tens of different motion clips and pose instances are usually located in different places, which can make it easier for classifier to learn by frame location instead of real pose state.

At the current stage we have not decided to add pose dynamics attributes such as velocities and accelerations.

To reduce computational complexity of machine learning methods we have prepared test set with only 2000 randomly selected pose frames.

3 Classification

First we have tested supervised learning methods based on the raw data containing all 56 relative attributes. We have used cross validation for to split our test set into the train and test parts and focused on the classifier efficiency, meaning percentage of correctly classified poses of the test sets.

We have chose following classifiers:

- Naive Bayes [10] with normal and kernel based density estimator,
- kNN [10] with number of analyzed nearest neighbors ranging from 1 to 10,
- Random Forest [10] with various number of features,
- MultiLayer Perceptron [10] with various numbers of hidden layer neurons and epochs plus several different learning rates.

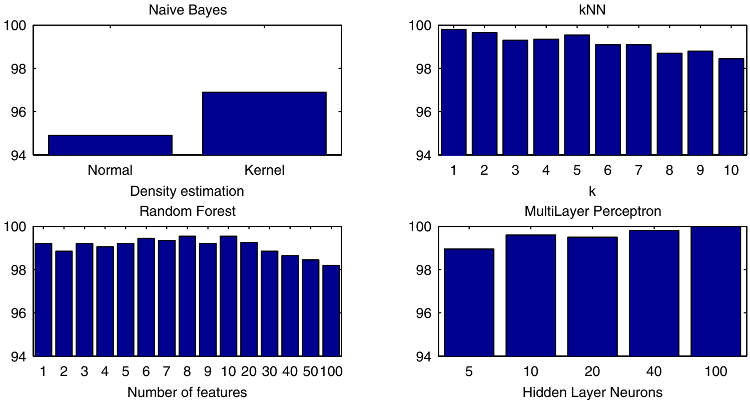


Fig. 2. Classification results

The efficiencies of all classifiers are over 95 percent and in best case of neural network classifier it comes 100 percent. All results are presented in Fig. 2.

For Naive Bayes, which achieved worst efficiencies, there is significant difference for normal and kernel based density estimator. The advantage of kernel based one probably means that the assumption of normal distribution of pose attributes is not so accurate, but the other hand, almost 95 percent efficiency does not deny normal distributions.

KNN classifier achieved very good results. There is opposite relation of number of the analyzed nearest neighbors and efficiency - the more neighbors the worse results. It is probably so because of the nature of our dataset, which in a few untypical feature space regions has weak representation of given pose type frames. The nearest neighbor classifier is best fitted to the train dataset regardless of that representation. In spite of that, efficiency of 10NN classifier is still acceptable and only a little bit worse than 1NN.

There is a slightly noticeable influence of number of features of random forest classifier, but the differences are not remarkable and all results are satisfactory.

Globally the best results are achieved with neural network classifier. The results depend proportionally on the complexity of the network - the greater complexity, the better results, but even five hidden layer neurons give excellent efficiency over 99 percent. We think it is so because of the above mentioned weak representation, which could be better approximated with more complicated networks.

4 Dimensionality Reduction

We have applied dimensionality reduction methods to reduce pose descriptions. On the basis of the reduced feature space, we have tested supervised learning methods and compared the results with raw data.

We have used and compared linear Principal Components Analysis [10] and nonlinear Kernel Principal Components Analysis [9] and tested nonlinearity of the feature space.

We have chosen radial kernel function $K(x, y) = e^{\frac{\|x-y\|^2}{2\sigma^2}}$ [1], with different sigma values and Eucalidean metric calculated on normalized and raw feature space.

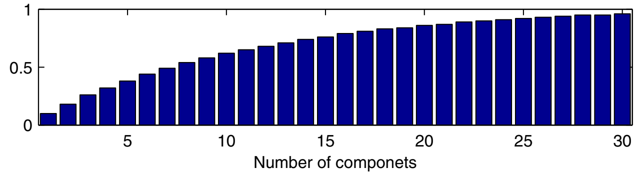


Fig. 3. PCA Variance Cover

Variance cover of PCA components shows that there is no short description which stores most of dataset variance. Three components have only 26 percent of global variance, and 90 percent receives just more than twenty componets.

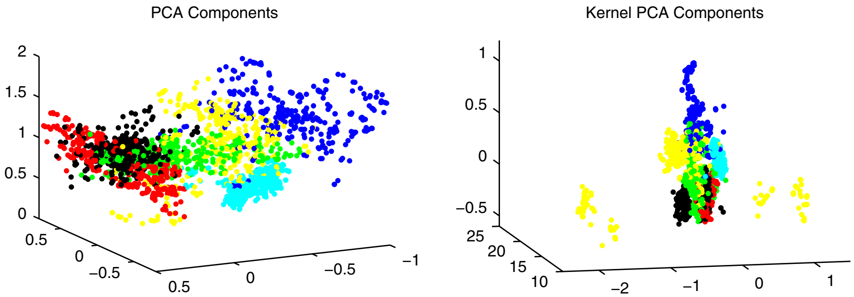


Fig. 4. Reduced feature space. Poses: blue-sitting, red-standing, green-jumping, yellow-climbing, black-walking, cyan-running.

Table 1. Example Confusion Matrices. 3D PCA feature space, Naive Bayes and Random Forest classifiers

	sit	sta	jum	cli	wal	run
sitting	283	3	0	47	1	0
standing	1	97	5	2	130	0
jumping	18	17	232	8	55	3
climbing	30	15	60	212	3	13
walking	0	62	61	1	209	0
running	0	2	8	0	0	323

	sit	sta	jum	cli	wal	run
sitting	312	0	1	19	1	1
standing	0	296	11	3	23	1
jumping	1	18	267	14	31	2
climbing	24	13	31	261	2	2
walking	1	41	34	2	254	1
running	0	0	7	2	1	323

In the Fig. 4 we have visualized the first three components of PCA and Kernel PCA in reduced 3D feature space and in Tab. 1 we have presented example confusion matrices for this 3D PCA feature space achieved with Naive Bayes and Random Forest classifiers. For Kernel PCA we have used kernel function parameters for which we have obtained best classification results. We can notice general pose classes boundaries, but there is no accurate simple distinction between them. Especially poses standing and walking are mixed together. It happens so because in slow walking there are some phases which look very similar to standing and three values are insufficient to distinguish them. In Kernel PCA some climbing poses are placed far from the rest of instances. There are probably poses with largely leaning forward, which produce large values of distance metric to the rest of posed and has an impact on the kernel function values.

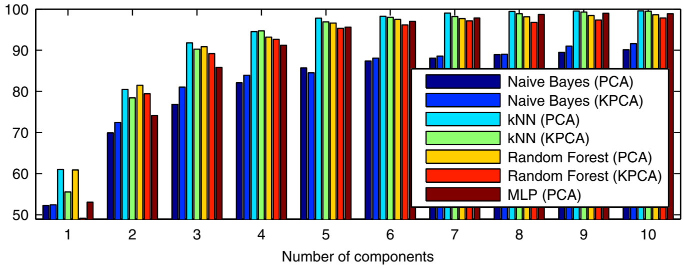


Fig. 5. PCA and Kernel PCA classification results

Fig. 5 shows aggregated classification results obtained by classifiers for PCA and Kernel PCA reduced feature space. We have chosen results for best parameters of each classifier and in the case of Kernel PCA best pair of sets of parameters classifier and kernel function.

There is a similarity to raw feature space. The worst is Naive Bayes and kernel density estimator is a bit better than normal. For kNN the best one is 1NN, but the variations are not remarkable. The only difference is that the multilayer perceptron is not better than others and does not need such complexed structure to obtain optimal efficiency. Globally the best is 1NN, but Random Forest and neural network are almost the same. Acceptable results with efficiency over 90 percent need at least three dimensional features space, for 95 percent five dimensional is required, but excellent 99 percent needs only seven dimensional. There are no remarkable deferences of PCA and Kernel PCA, except for one dimensional feature space which promotes PCA. Kernel PCA is a little bit better for Naive Bayes but a bit worse for kNN and Random Forest.

Kernel function has great impact on the classification. In most cases distance metric calculated on the normalized feature space is promoted. There is no noticeable general dependency as regards sigma parameter, it differs in given cases.

We also built classifier based on the first components of the Kernel PCA trained on the datasets with pose frames of only single pose type. Kernel function depends on the similarity of its arguments, the more similar the greater the value. Thus, sum of kernel function values calculated against the same pose type could give greater value than against other pose type. We have decided to assign pose to the class with maximum value of first Kernel PCA component trained on the given class instances.

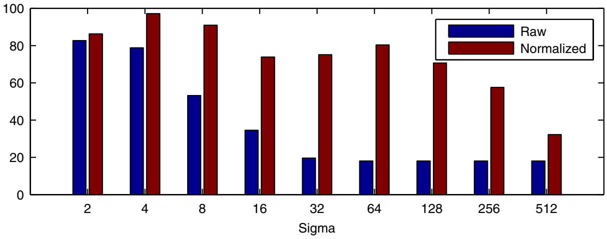


Fig. 6. Kernel PCA classifier results for distance metric calculated on the raw and normalized feature space

We have obtained over 97 percent of classifier efficiency for the best case. Regardless of sigma values, the distance metric based on the normalized feature space gives better results. The choice of analyzed Kernel PCA component is disputable and there is possible area to improve the results.

5 Conclusion

We have evaluated supervised learning techniques for detection of pose type, based only on the location of body markers. We have prepared test database of 2000 pose frames and six different pose types. We have chosen four different classifiers and tested them on wide range of parameters. The results are very promising, we have obtained even 100 percent of classifier efficiency for the case of multilayer percptron with complexed structure. However, it could be an overtrained network, ideally fitted to the train dataset. Although train and test datasets are disjoint, they come from the same database and have some kind of dependency. In fact there is no possibility to prepare dataset with unique cover of possible class regions for the data described with 56 attributes. A single pose type can be represented by a very large number of different frames, and some attributes may have no significance, like the position of the hands in the sitting pose.

Dimensionality reduction techniques preserve global information of the pose state. Three-dimensional feature space is sufficient to notice the inaccurate boundaries between pose types, but better results require more dimensions. For ten-dimensional space the results are rated on the level of 99 percent, which is only slightly worse than for the full 56 attributes. Dimensionality reduction generalizes feature space ; thus, it diminishes focusing on the pose details and strict fitting to train dataset. We think that results are more reliable.

Our experiments are only introductory stage to real applications, which is more challenging task, because of the above mentioned train set representation and more pose types.

Our final conclusion is that supervised machine learning techniques are able to recognize pose types.

Acknowledgment

This paper has been supported by the project „System with a library of modules for advanced analysis and an interactive synthesis of human motion" co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme - Priority Axis 1. Research and development of modern technologies, measure 1.3.1 Development projects.

References

1. Boser, B.E., Guyon, I.M., Vapnik, V.: A training algorithms for optimal margin classifiers. In: Fifth Annual Workshop on Computational Learning Theory, Pittsburgh (1992)
2. Carnegie-Mellon Mocap Database (2010), `http://mocap.cs.cmu.edu`
3. Johnson, M.: Exploiting Quaternions to Support Expressive Interactive Character Motion. PhD thesis, Massachusetts Institute of Technology (2003)
4. Kovar, L., Gleicher, M.: Flexible automatic motion blending with registration curves. In: Proc. 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 214–224 (2003)
5. Kovar, L., Gleicher, M., Pighin, F.: Motion graphs. ACM Trans. Graph., 73–482 (2002)
6. Lee, M.W., Cohen, I.: A Model-Based Approach for Estimating Human 3D Poses in Static Images. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(6) (2006)
7. Müller, M., Röder, T.: 00 A Relational Approach to Content-based Analysis of Motion Capture Data. In: Computational Imaging and Vision, ch. 20, vol. 36, pp. 477–506 (2007)
8. Roder, T.: Similarity, Retrieval, and Classification of Motion Capture Data. PhD thesis, Massachusetts Institute of Technology (2006)
9. Schoelkopf, B., Smola, A., Mueller, K.-R.: Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Technical Report No. 44, Max-Planck-Institut fuer biologische Kybernetik (1996)
10. Witten, I., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco (2005)

Region Covariance Matrix-Based Object Tracking with Occlusions Handling

Ivar Austvoll[1] and Bogdan Kwolek[2]

[1] University of Stavanger, N-4036 Stavanger, Norway
ivar.austvoll@uis.no
[2] Rzeszów University of Technology, 35-959 Rzeszów, Poland
bkwolek@prz.edu.pl

Abstract. This work proposes an optical-flow based feature tracking that is combined with region covariance matrix for dealing with tracking of an object undergoing considerable occlusions. The object is tracked using a set of key-points. The key-points are tracked via a computationally inexpensive optical flow algorithm. If the occlusion of the feature is detected the algorithm calculates the covariance matrix inside a region, which is located at the feature's position just before the occlusion. The region covariance matrix is then used to detect the ending of the feature occlusion. This is achieved via comparing the covariance matrix based similarity measures in some window surrounding the occluded key-point. The outliers that arise in the optical flow at the boundary of the objects are excluded using RANSAC and affine transformation. Experimental results that were obtained on freely available image sequences show the feasibility of our approach to perform tracking of objects undergoing considerable occlusions. The resulting algorithm can cope with occlusions of faces as well as objects of similar colors and shapes.

1 Introduction

Tracking an object in a sequence of images is currently utilized in many computer vision applications. The goal of visual tracking is to locate a region in each image that matches an appearance of a target object. The algorithms for visual object tracking can be divided broadly into two categories, namely: feature-based and visual-model/template-based [1]. Feature-based methods track an object through tracking a collection of local features such as corners [2]. The second group of methods achieves the object tracking through matching a template or a model to the input image [3].

In typical scenarios, interactions between moving objects result in partial or significant occlusions, making the object tracking a highly challenging problem. Various systems and methods have been proposed to handle object tracking in complex crowded scenes with the occlusions arising in the object tracking [4]. Multiple cameras are often used to cope with occlusions [5]. In most conventional multi-camera systems the targets are represented as a collection of blobs in 3D space, which are tracked over time. This requires finding the corresponding blobs across

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 201–208, 2010.

multiple images as well as assigning 2D blobs to the 3D blobs. However, using a multi-camera system in many applications may be impractical. Therefore, stereo cameras are often used to perform object tracking in such circumstances [6]. However, conventional stereo cameras usually do not provide useful and reliable depth estimates in occluded regions, particularly when they are texture-less.

The object tracking is often achieved using a single camera. However, one fundamental limitation of using one camera in the tracking of objects is dealing with object occlusions. In single-camera methods, occlusion can be identified through prediction of the object location or on a per-pixel basis. Kalman filtering or particle filtering [7] can be used to predict the positions of objects during occlusions. Methods relying on per-pixel representation often use templates to represent objects. The underlying main assumption behind template matching is that the appearance of the template remains almost the same throughout the entire image sequence. Hence, handling occlusions is not an easy task in such an approach [3]. Babenko *et. al* [8] recently proposed an online multiple instance learning algorithm to achieve robust object tracking under occlusion. However, to achieve long-term object tracking, a persistent tracker must cope with occlusions as well as must be able to reacquire the object in case of considerable occlusions.

Despite the above advances, in many situations the existing algorithms do not have satisfactory tracking robustness, especially when there is a large amount of occlusion between two or more objects. Therefore, a highly efficient occlusion handling scheme, which could lead to a considerable improvement of the tracking performance, even when there is a large amount of occlusion between two or more objects is needed. Our approach to cope with considerable occlusions is to construct a region covariance matrix in the surround of the feature just before occlusion and then to employ such a descriptor to detect the ending of the occlusion. The motivation behind such an approach is that the region covariance matrix is a strong and robust indicator for point-to-point correspondence. It is a powerful descriptor that encodes the variance of the channels, such as red, green, blue, gradient, etc., their correlations with each other, and the spatial layout [9]. Moreover, variations in illumination as well as in pose or viewpoint do not affect the covariance considerably. Through the use of such a robust region descriptor a tracked feature can be detected and recognized again after loosing it.

The features are tracked using the optical flow. In optical flow-based feature tracking the significant errors might occur on the occluding boundary. In [10] it has been shown that features belonging to the same object have correlated behavior, whereas features belonging to different objects show evidence of more uncorrelated and independent behavior. Motivated by this observation we assume that key-points in previous and current images are related by an affine transformation, and we then try several combinations of 3 points in a RANSAC framework to exclude outliers, i.e. features that do not move consistently with the inliers. This helps us to detect the occlusion, and more importantly, the features do not undergo undesirable shifting through the appearance changes at the boundaries where occlusions take place.

In Section 2 we present a feature-based object tracking and start with a discussion of feature detection and optical flow estimation. Thereafter we present how the consistency of matches is handled in our approach. Then we outline covariance matrix based region descriptor as well as present our algorithm. Section 3 is devoted to demonstration of experimental results. We end the paper with conclusions.

2 Feature-Based Object Tracking

2.1 Feature Detection and Optical Flow Estimation

Good detectors of features are very important for object tracking [2]. Several feature detectors and descriptors have been proposed and evaluated in the literature [11][12]. Recent research [12] has demonstrated that the repeatability of the key-point detectors deteriorates with change of the viewpoint. The work mentioned above has also demonstrated that no key-point detector performs well in case of considerable view changes. Taking this into account we utilize the Harris corner detector [13] in our algorithm. Another rationale of our choice is that the Harris corner detector has relatively low computational cost when compared to the SIFT algorithm.

The inter-frame translations of the key-points are determined by the Lucas-Kanade optical flow algorithm [14]. This method is still one of the best methods for two-frame motion estimation. The advantage of the method is that the features can be tracked with low computational cost and therefore it is utilized in our algorithm.

2.2 Consistency of Matches

RANSAC (RANdom SAmple Consensus) is a robust method to estimate parameters of a mathematical model from a set of data contaminated by considerable amounts of outliers [15]. The percentage of outliers which can be handled by RANSAC can be larger than 50% of the entire data set. The RANSAC algorithm consists of two steps, which are repeated in an iterative hypothesis-and-test fashion. In the hypothesis step, minimal sample sets are randomly chosen from the input dataset and then the model parameters are estimated using only elements from such sets. In the second step, the RANSAC tests whose elements of the entire dataset are consistent with a model, which is instantiated with the parameters from the first step. The steps mentioned above are repeated a fixed number of times. Each time, either a refined model is produced or the model is declined because too few points are classified as inliers.

In our approach we employ RANSAC to find the largest set of matches consistent with an affine transformation. The affine transformation is given by the following equation:

$$\begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{1}$$

where $[x'_i \;\; y'_i \;\; 1]$ is the matched feature location in the current image and $[x_i \;\; y_i \;\; 1]$ is the matched feature location in the previous image. A sum of square errors is minimized to estimate the affine transform parameters $h = [h_1 \;\; h_2 \;\; h_3 \;\; h_4 \;\; h_5 \;\; h_6]$ according to the following equation:

$$\min_h \sum_i (h_1 x_i + h_2 y_i + h_3 - x'_i)^2 + (h_4 x_i + h_5 y_i + h_6 - y'_i)^2 \qquad (2)$$

The solution to the optimization problem (2) is given by the following equation:

$$h = A^{-1} b \qquad (3)$$

where

$$A = \sum \begin{bmatrix} x_i^2 & x_i y_i & x_i \\ x_i y_i & y_i^2 & y_i \\ x_i & y_i & 1 \end{bmatrix} \qquad (4)$$

and

$$b = \sum [x'_i x_i \;\; x'_i y_i \;\; x'_i \;\; y'_i x_i \;\; y'_i y_i \;\; y'_i]^T \qquad (5)$$

2.3 Covariance Matrix Based Region Descriptor

Recently, in [16] an elegant and simple solution to integrate multiple image features has been proposed. It is based on the covariance matrix. Using a covariance matrix (CM) as a region descriptor has many advantages, namely: 1) CM indicates both spatial and statistical properties of the objects; 2) it provides an elegant means to combine multiple modalities and features; 3) it is capable of relating regions of different sizes.

Let $\mathbf{I}$ be an image of size $W \times H$. At each pixel location $\mathbf{x} = [x, y]^T$ we can extract d features such as intensity, gradient, color, etc. Using such a feature set we can construct a $W \times H \times d$ feature image $\mathbf{H}$. Given a rectangular window R we can then compute the covariance matrix $\mathbf{c}_R$ of the features according to the following equation:

$$\mathbf{c}_R = \frac{1}{|R| - 1} \sum_{\mathbf{x} \in R} (\mathbf{H} - \mathbf{m}_R)(\mathbf{H} - \mathbf{m}_R)^T \qquad (6)$$

where $\mathbf{m}_R = \frac{1}{|R|} \sum_{\mathbf{x} \in R} \mathbf{H}(\mathbf{x})$ denotes the vector of means of corresponding features for the pixels in region R, and $|R|$ stands for the size of region R. The diagonal entries in such a covariance matrix express the variance of each feature and the off-diagonal entries indicate their mutual correlations. The covariance matrix is a very informative region descriptor because it encodes information about the variance of features, their correlations witch each other, and spatial layout. It can be computed efficiently through the use of integral images in a way that has been shown in [9]. To measure the dissimilarity between the covariance matrixes $\mathbf{c}_1$ and $\mathbf{c}_2$ we employed the following distance [16]:

$$\rho(\mathbf{c}_1, \mathbf{c}_2) = \sqrt{\sum_{i=1}^{|R|} \ln^2 \lambda_i(\mathbf{c}_1, \mathbf{c}_2)} \qquad (7)$$

where $\{\lambda_i(\mathbf{c}_1, \mathbf{c}_2)\}_{i=1,\ldots,|R|}$ are the generalized eigenvalues of $\mathbf{c}_1$ and $\mathbf{c}_2$, which are calculated on the basis $\lambda_i \mathbf{c}_1 \mathbf{x}_i - \mathbf{c}_2 \mathbf{x}_i$, where $\mathbf{x}_i \neq 0$ are the generalized eigenvectors. Another possibility to measure the similarity between covariance matrices is to use Log-Euclidean metrics [17].

2.4 The Algorithm

When a new frame is available, after detecting Harris corners, the optical flow is estimated to determine the current location of key-points. At this stage we calculate the quality of the features [2] in order to verify if they are still trackable and have not drifted away from original targets. Through monitoring the features' quality we verify whether each feature is occluded or not. In case of an occlusion we calculate the region covariance matrix at feature's location before the occlusion and afterwards we finally decide if the occlusion takes place. For the non-occluded features we apply RANSAC with the affine model in order to determine the outliers, i.e. features that move inconsistently according to the best affine model. For such features we compute region covariance matrixes and add the features to the set of occluded features. This way we suppress the motion errors that arise at the boundary of the occlusions. In subsequent frames for each occluded feature we execute a test if the occlusion is finished. Given the feature location before the occlusion we perform greedy search in a window surrounding such a location for the best similarity of region based covariance descriptors. If the best distance between covariance matrixes is below the threshold we start the tracking of the feature. In [18] the RANSAC algorithm is used to identify consistent subsets of correspondences and obtain a better homography. Our work differs from the mentioned work in that we focus on handling the occlusions of the object undergoing tracking with the support of the RANSAC.

3 Experimental Results

We validated the algorithm by tracking a face, which undergoes considerable occlusions[⋆]. Although almost the whole face was occluded our tracker successfully tracks the face, see images in upper row of Fig. 1, as well as detects it after the occlusion, see images in bottom row. As we can observe the algorithm is able to reacquire the object despite similar colors as well as textures of both objects. The above mentioned similarity of both objects leads to sporadic misdetections of the feature's occlusion as it can be seen in frame #7, where some features are located at the occluding hand. Through the use of RANSAC built on an affine motion model, such feature drifting is eliminated quickly as can be observed in frame #9. In frame #13 we can notice that the algorithm reacquired most of the features. The location of the features is consistent with their location before occlusion, see frame #5. In the next frames of the discussed sequence we can perceive the behavior of the algorithm after redetection of the features and during

[⋆] Thanks Dr. Birchfield for this sequence, obtained from
`http://robotics.stanford.edu/~birch/headtracker`

the subsequent occlusion. As we can see at frame #26 the number of redetected features is sufficient to continue the tracking of the face. The locations of the redetected features are consistent with the location before the second occlusion, see frame #13, as well as with initial feature locations, see frame #5.

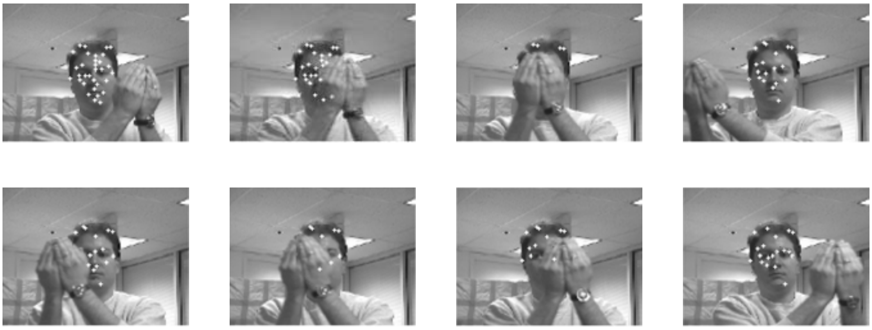


Fig. 1. Object tracking during considerable occlusion. Frames #5, 7, 9, 13 (upper row), #20, 21, 23, 26 (bottom row).

The experimental results shown in Fig. 2 demonstrate that the RANSAC algorithm allows us to obtain far better tracking results. In particular, as experimental results show, without RANSAC some features undergo undesirable shifts and in consequence the number of the reacquired features is somewhat smaller, see frame #26. Even more, as we can see at the mentioned image, some features can be located in the wrong objects.

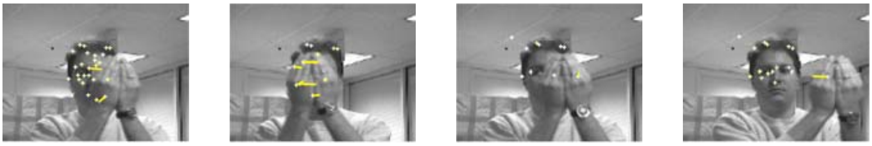


Fig. 2. Key-point-based object tracking without RANSAC. Frames #7, 9, 23, 26.

Figure 3 demonstrates a scene in which a face undergoing tracking is temporally occluded by another one. The occluded face moves slightly during the occlusion and in consequence the number of the reacquired features is something smaller. The extension of the algorithm about Procrustes analysis [19] to compute the similarity transform (translation, rotation and scale) between two sets of visible features and then to transform the occluded features is straightforward.

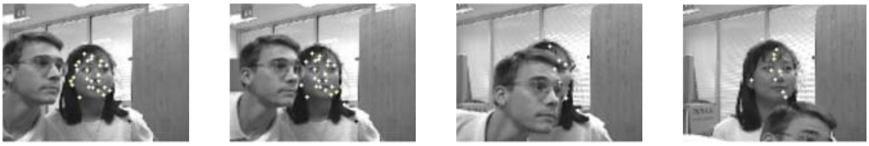


Fig. 3. Tracking a face that undergoes occlusion by an another face. Frames #420, 422, 434, 449.

The algorithm has been implemented in Matlab. The above described experiments were done on color images of size 128 × 96. The recovery of the occluded feature is done through the greedy search for the best similarity of the region covariance matrixes and then comparing it with a threshold value. The searching is realized in a window of size 5 × 5 centered on feature's position just before the occlusion. The region covariance is built in a windows of size 6 × 6 using feature location, R, G, B color values and first and second image derivatives.

4 Conclusions

To persistently track an object in long image sequences the algorithm must cope with considerable occlusions. Since the existing algorithms can not perform well under considerable occlusions, we propose an algorithm that employs region covariance descriptors to reacquire the occluded features. We demonstrated experimentally that such a descriptor is very useful in recovering the features. The RANSAC algorithm helps considerably in detecting the occlusions as well as allows us to exclude outliers arising at the boundary between moving objects.

Acknowledgment

Portions of this work were completed by B. Kwolek in the Faculty of Science and Technology at the University of Stavanger during a research stay, which has been supported within the EEA Financial Mechanism and the Norwegian Financial Mechanism.

References

1. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Comput. Surv. 38, 13 (2006)
2. Shi, J., Tomasi, C.: Good features to track. In: Proc. of CVPR, pp. 593–600 (1994)
3. Schreiber, D.: Robust template tracking with drift correction. Pattern Recogn. Lett. 28, 1483–1491 (2007)
4. Gabriel, P.F., Verly, J.G., Piater, J.H., Genon, A.: The state of the art in multiple object tracking under occlusion in video sequences. In: Int. Conf. on Advanced Concepts for Intelligent Vision Systems, pp. 166–173 (2003)

5. Khan, S., Shah, M.: A multiview approach to tracking people in crowded scenes using a planar homography constraint. In: Proc. of the 10th European Conf. on Computer Vision, Graz, Austria, pp. 133–146 (2006)
6. Darrell, T., Gordon, G., Harville, M., Woodfill, J.: Integrated person tracking using stereo, color, and pattern detection. Int. J. Comput. Vision 37, 175–185 (2000)
7. Isard, M., Blake, A.: Condensation - conditional density propagation for visual tracking. Int. J. of Computer Vision 29, 5–28 (1998)
8. Babenko, B., Yang, M.H., Belongie, S.: Visual tracking with online multiple instance learning. In: IEEE Comp. Society Conf. on Computer Vision and Pattern Recognition, Miami, Florida, USA, pp. 983–990 (2009)
9. Tuzel, O., Porikli, F., Meer, P.: Region covariance: A fast descriptor for detection and classification. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 589–600. Springer, Heidelberg (2006)
10. Ramanan, D., Forsyth, D.A.: Using temporal coherence to build models of animals. In: Proc. of the Ninth IEEE Int. Conf. on Computer Vision, Washington, DC, USA, pp. 338–345. IEEE Computer Society, Los Alamitos (2003)
11. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. Journal of Computer Vision 60, 91–110 (2004)
12. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. 27, 1615–1630 (2005)
13. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. of Fourth Alvey Vision Conference, Manchester, UK, pp. 147–151 (1988)
14. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proc. Int. Joint Conf. on Artificial Intell., pp. 674–679 (1981)
15. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. ACM Commun. 24, 381–395 (1981)
16. Porikli, F., Tuzel, O., Meer, P.: Covariance tracking using model update based on lie algebra. In: Proc. Int. Conf. on Comp. Vision and Pattern Recognition, vol. 1, pp. 728–735 (2006)
17. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Fast and simple calculus on tensors in the log-euclidean framework. In: Int. Conf. on Medical Image Computing and Computer Assisted Intervention, pp. 115–122 (2005)
18. Okuma, K., Little, J.J., Lowe, D.G.: Automatic rectification of long image sequences. In: Asian Conference on Computer Vision (ACCV), Jeju Island, Korea (2004)
19. Mardia, K., Dryden, I.: Statistical Shape Analysis. Wiley, Chichester (1998)

Minimalist AdaBoost for Blemish Identification in Potatoes

Michael Barnes, Grzegorz Cielniak, and Tom Duckett

School of Computer Science, University of Lincoln, Lincoln LN6, 7TS, UK
{mbarnes,gcielniak,tduckett}@lincoln.ac.uk

Abstract. We present a multi-class solution based on minimalist AdaBoost for identifying blemishes present in visual images of potatoes. Using training examples we use Real AdaBoost to first reduce the feature set by selecting five features for each class, then train binary classifiers for each class, classifying each testing example according to the binary classifier with the highest certainty. Against hand-drawn ground truth data we achieve a pixel match of 83% accuracy in white potatoes and 82% in red potatoes. For the task of identifying which blemishes are present in each potato within typical industry defined criteria (10% coverage) we achieve accuracy rates of 93% and 94%, respectively.

1 Introduction

In this paper we present a multi-class minimalist AdaBoost solution to identifying blemish types in images of potatoes. Potatoes are an important part of the food supply in much of the world and the biggest factor in their market price is appearance. A typical human grader can make a quality classification for a small baking potato every 0.45 seconds, thus there is a need to consider both speed and accuracy in an automatic classifier. The system developed should be trainable, so that it can work with different varieties of potatoes and variations in seasons, lighting conditions, etc. A human expert is required to mark up areas of blemishes by type in a set of training images. After training, the system should be able to classify individual pixels by their blemish type, and hence grade whole potatoes according to typical industry criteria, in this case 10% coverage.

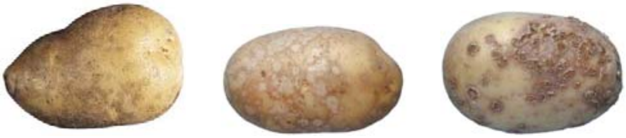
1.1 Potato Blemishes

There are a number of diseases affecting potato tubers that, although superficial and generally of little or no health consequence to humans, strongly and negatively influence consumer choice. These conditions present a variety of differently coloured, sized and textured symptoms on the skin surface. Such diverse visual information provides us with a rich source of indicators that can be used for training an automatic blemish detector.

These blemishes include black dot, a mostly speckled blemish with a soft texture ranging from brown speckles to a silvery area with black speckles. Silver

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 209–216, 2010.

scurf is a fungal infection which manifests as a silvery area, often with a brown perimeter and black speckles. Powdery scab and common scab both manifest primarily as areas where the skin splits open along straight lines presenting an often star-like shape. Other forms of blemish include physiological blemishes, e.g. greening and sprouting. See Fig. 1 for examples of these blemish classes.



(a) A potato blemished by black dot.

(b) A potato blemished by silver scurf.

(c) Potato blemished by powdery scab.

Fig. 1. Examples of the potato blemish types used in our experiments

The identification of blemishes is important for a number of reasons. Supermarkets will usually be more tolerant of some blemishes than others, for instance greening, of the bright "field green" variety, is more of an issue than silver scurf, since the latter can be peeled off. Secondly, there is an interest in the possibility of using such a system for epidemiology purposes, to determine which farms are growing more potatoes with certain blemishes in order to coordinate efforts to provide treatments such as fungicides to the regions most afflicted.

1.2 Related Work

In typical machine vision systems for quality analysis of food products, there are several major steps: after pre-processing (e.g. to segment the object of interest from the background), features are extracted that summarise important qualities of the object, then a pattern recognition system is used to categorise the input data. For example, [1] developed methods to distinguish between blemishes in apples and healthy apples with visible stem or calyx. In [2], HSV-colour and texture histograms are used to classify different types of fruit and vegetables.

In the area of machine vision for potatoes, [3] used Fourier harmonics to describe the shapes of potatoes. Custom lighting equipment is used in [4] to project light at a variety of different wavelengths to demonstrate the different reflective properties of specific blemishes at each wavelength. [5] used the HSI colour space to identify green potatoes as well as yellow and green apples. [6] graded potatoes by size and shape. In [7] a system uses green levels to detect green defects (greening and sprouting) in potatoes. In [8] mis-shapen potatoes are detected by comparing the local rate of change in radius of a potato as well as detecting sprouting by comparing the green colour channel with the intensity.

A major limitation of typical systems is that the set of features for pattern recognition has to be designed by the system engineer to work with a specific configuration of produce, imaging system and operating conditions. Such systems typically do not generalise well to other configurations, where the required features may well differ from those used to design the original system. In previous work, [9] we used a single class method using AdaBoost [10] to both select good features for a particular pattern recognition task and then to select weak classifiers based on the features chosen. In this paper we extend the approach to identify the different types of blemish described in Section 1.1.

2 Materials and Methods

2.1 Image Acquisition

The experimental data for this system, consisting of images of potatoes, were acquired using a colour camera (Sony DSLR-A350K) fixed above the tubers which in turn were placed on a white board. The camera was set at a distance of 60cm from the camera objective to the base on which the subjects were placed, with a focal length of 70mm and an aperture setting of F22. The resolution of the images was 1536×1024 pixels. To reduce the effects of shadows and changing light conditions the potatoes were placed inside a white cylinder with daylight bulbs placed around the top.

2.2 Ground Truth

There were two sets of data collected for white and red potatoes respectively, including potatoes affected by different blemishes. For our classifier we combined greening and sprouting into one class as well as combining powdery and common scab into a single class, since the latter pair are very difficult to tell apart [11]. We are left with five classes; black dot, silver scurf, scab, green and unblemished. The white potato data set contains 102 images including 19 images containing a single blemish type, 39 images with two distinct blemish types, 38 images with three blemish types and 6 images containing more than three blemish types. The most common blemishes were black dot and silver scurf, appearing in 69 and 53 images respectively, while the rarest were powdery scab, elephant hide and growth cracks, with no more than 3 images of each. The red potato data set contains 48 images including between 10 and 15 examples of each blemish class.

To train the classifiers and test their performance, the images need to be marked up by hand to provide the "ground truth" information indicating the correct class of each pixel. The mark up process begins with a semi-automatic method for background removal, using the Magic Wand tool in Adobe Photoshop to label the image region surrounding the potato. The potato area is then hand labelled by an industry expert into regions corresponding to blemish types and to non-blemish. It is not necessary to label all pixels in an image: some areas of high uncertainty or ambiguity are left unmarked, and these pixels are ignored during training of the classifier. Background pixels are also omitted from the subsequent calculations. An example of training and ground truth images can be seen in Figure 2.

2.3 Feature Extraction

The first step of the procedure is to extract image features that should indicate the type of blemish present in a potato image. The features include statistical summaries of the whole potato and square regions of interest centred on each pixel as well as the data of the pixel itself. The statistics used for each feature type and region were the mean, variance, skew, maximum and minimum values in the area. The proposed system uses the RGB colour space - the original colour format of the camera output. The system software was implemented in MATLAB.

The square regions used in our experiments were of size 33 × 33, 65 × 65, 97 × 97, 129 × 129 and 161 × 161, giving 5 regions in total, as well as the pixel.

Our system uses seven colour channels; raw RGB, normalised RGB and the grayscale intensity. From these channels we consider the following image properties:

Colour and grayscale intensity: Intensity is especially of relevance for dark blemishes, e.g. black scurf or skin spot, while the most obvious blemish to be detected by other colour channels would be greening.

Edge gradient: An edge detector determines the rate of change of pixel values in a given neighbourhood in a specific direction. Some blemishes tend to coincide with high rates of change, such as powdery scab when the skin splits. The Sobel edge detector was used in this case with a standard 3 × 3 kernel size.

Edge length: An extension of the edge gradient feature, the edge length is determined by using a thresholded Sobel edge detector. Pixels are first labelled as edge or non-edge, then edge pixels are used to form connected components. Each pixel within a component is then given a value equal to the number of pixels which form that component. Non-edge pixels are given a value of zero. Larger edge components tend to be found around scabs in particular.

Range: The range filter determines the maximum difference between pixel values in a given neighbourhood indicating the roughness of the texture. Higher values tend to correspond to rougher, potentially damaged areas of the image. The range filter was run on the same seven channels with a 5 × 5 neighbourhood.

In summary there are 7 colour channels × 4 feature types × 5 statistics making 140 features for each region and $7 \times 4 = 28$ features for the pixel itself. All these features are used as the candidate feature set. Since there are 5 different regions, this gives $140 \times 5 = 700$ features which, with additional 28 features for the pixel itself, gives us $F_c = 728$ candidate features in total. These features were used as the training input to our classifier.

2.4 AdaBoost

The AdaBoost algorithm [10] is used to build a classifier, which combines results from so-called "weak" classifiers (each a decision stump using one of the candidate features) into one "strong" classifier that performs better than any of the weak classifiers alone. The high performance of the final strong classifier is due to the emphasis put on the training examples which are most difficult to classify during the learning process. This method is called boosting. During training AdaBoost makes a number of passes, called rounds or iterations, through the training data. Each time it finds the next best feature to improve the number of correctly classified examples, prioritising those examples which were misclassified previously. In each pass one feature is selected and assigned a weight and a threshold to create a new weak classifier. The weak classifiers are then combined into a strong classifier wherein each weak classifier has a weighted vote in the classification of a given example.

Real AdaBoost [12] is a generalisation of this algorithm that provides a lower error rate by allowing weak classifiers to vote by their individual degree of certainty instead of simply voting "yes" or "no". It is the version used in our experiments, hereafter referred to simply as AdaBoost. Using MATLAB we have used the Real AdaBoost implementation within the GML AdaBoost Toolbox [13] for these experiments.

Minimalist AdaBoost [9] is a method of limiting the number of unique features used by Real AdaBoost while still allowing free selection of a larger number of weak classifiers from the whole set of candidate features. This is achieved by training on the full set of features until AdaBoost has selected weak classifiers using a chosen number of unique features, then continuing using only those selected features up to the chosen maximum number of weak classifiers. This can improve the speed of a system which has to extract features in real time, in tasks such as produce quality control.

To extend Minimalist AdaBoost to multi-class recognition problems, it was trained for each class with 10,000 positive examples and 10,000 negative examples, the negative examples being split equally between all other classes. First the classifier was used to select features for each class, then each binary classifier was trained on all selected features, allowing 40 weak classifiers. When classifying test data, the five different binary classifiers were used and their output certainties compared. The pixel was then classified according to the classifier which returned the highest certainty of a positive result for its class, or the lowest certainty of a negative result (a winner-takes-all approach).

3 Results and Discussion

3.1 Training and Testing

To begin, the images were split into training and testing sets. These were selected to ensure that a representative number of examples of every blemish was present in both sets and also that those potatoes which had been photographed twice, front and back, did not appear in both sets. From this data, 10,000 examples of each of the 5 classes were extracted, spread out within all training images.

Using minimalist AdaBoost with a winner-takes-all approach, each example (an individual pixel in an image) was classified with the binary classifiers and assigned to the class with the highest certainty.

3.2 Selected Features

In many cases the top features selected for each class had an intuitive explanation for why they were chosen.

In white potatoes, the top feature selected for scabs was the variance of the grayscale intensity range filter over the smallest area, representing how much the texture changes in the immediate vicinity of this blemish. The other four were all related to edge detectors. The top feature selected for green blemishes was the minimum value of the green colour channel in the 65x65 region. Two interesting parallels occur between silver scurf and black dot; the presence of the maximum red for the smallest area and the presence of the variance of a normalised colour followed immediately by the maximum value of the same, non-normalised colour, which may be representative of their similarity as blemishes. For the good potato class, the top five selected features all relate to either the red or normalised red colour channels.

In red potatoes, the top feature selected for scabs is the maximum level of the red channel over the smallest area, possibly because there is more of a colour difference between scab and red skin than white, although three of the next four features are edge related. Green begins with two features relating to the intensity channel, which may be because field greening of red potatoes is closer to black than green. Features selected for black dot have some correlation with those on white potatoes, notable the first feature relates to the blue channel in the largest region, compared to the normalised blue channel in the largest region for white potatoes. Also the maximum red value in the smallest region is used again. For silver scurf the variance and skew of the edge detector for the normalised blue colour channel over the largest region are selected as well as two features for the green colour channel. For good potato skin, the most obvious difference from white potatoes is that only one of the five selected features relates to the red channel. Edge length is selected twice, once for green and once for normalised green.

3.3 Classification Performance

Table 1 shows a comparison between the classifier output and ground truth data. The main disagreements are fairly consistent. In white potatoes the biggest confusion is between silver scurf and black dot, which are known to look very similar at times, even confusing human experts. Figure 2 shows a very good example of the other common disagreement between the classifier and ground truth, which is the exact extent of a speckled blemish, in this case black dot.

It is likely that potential users of this system will be more interested in the performance of this classifier on a potato-wise rather than pixel-wise basis, since they will either be grading whole potatoes or monitoring, for instance, the number of potatoes afflicted by specific blemishes in certain fields, with reference either to location or to experiments such as pesticide trials. Potato-wise performance is shown in Table 2, with each class being given a positive result if it is detected in more than 10% of the pixel-wise results for that image, so 10% scab coverage was either wrongly detected or missed in a total of 5 potatoes.

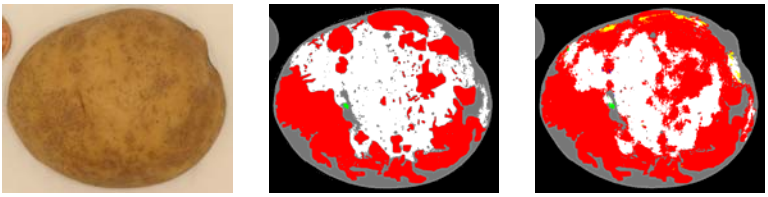
Table 1. Pixel-wise confusion matrices, showing the ground truth classes in rows and the classifier output classes in columns. (Bd = black dot, Ss = silver scurf, Sc = scab, Gr = green, un = unblemished).

(a) White potatoes

	Bd	Ss	Sc	Gr	Un
Bd	**80%**	9%	1%	0%	9%
Ss	8%	**82%**	5%	0%	4%
Sc	4%	7%	**87%**	0%	2%
Gr	1%	0%	0%	**96%**	3%
Un	9%	3%	1%	2%	**85%**

(b) Red potatoes

	Bd	Ss	Sc	Gr	Un
Bd	**63%**	5%	2%	5%	25%
Ss	6%	**72%**	3%	3%	17%
Sc	1%	1%	**97%**	1%	2%
Gr	1%	2%	1%	**92%**	4%
Un	8%	5%	3%	1%	**83%**



(a) Original potato

(b) Ground truth done by hand

(c) Classifier output

Fig. 2. A potato classified by the system, showing the similarity between classifier output and ground truth, with differences in the exact boundaries of blemished areas. Red = black dot, green = green, white = unblemished, grey+black = untested

Table 2. Potato-wise accuracy for detecting significant (10%) coverage per class

	White Potatoes	Red Potatoes
Black dot	90%	92%
Silver Scurf	88%	90%
Scab	95%	100%
Green	98%	96%
Unblemished	95%	96%
Overall	93%	95%
Blemish	93%	94%

The exact coverage that is considered as a major blemish varies by customer, typically in the range of 1% - 10% of surface area. Using 10% as our threshold for individual images we achieved a potato-wise classification accuracy of 93% and 94% for white and red potatoes, respectively.

4 Conclusions and Further Work

We have presented a trainable classifier using a multi-class adaptation of minimalist AdaBoost which detects blemishes in potatoes with 93-94% accuracy.

The biggest concern we identified was the disagreement as to the exact locations of blemishes between the hand-drawn ground truth and the pixel-wise classification. Some of these disagreements are due to the inaccuracy of hand-drawing as a means of ground truthing. It is possible that this could be improved with an assisted ground truthing method which might include some automatic classification during the markup process.

So far the presented method does not consider the 3d shape of potatoes, whereby a blemish seems smaller toward the edge of a potato since it is at a steeper angle to the camera. This challenge could be approached in several ways, either by increasing the weight given to blemishes near the potato's edge, using a mathematical model or a 3d sensor to decide the weighting, or by taking a number of images and averaging the detected blemish areas.

References

1. Unay, D., Gosselin, B.: Stem and calyx recognition on 'jonagold' apples by pattern recognition. Journal of Food Engineering 78, 597–605 (2006)
2. Bolle, R., Connell, J., Haas, N., Mohan, R., Taubin, G.: Veggievision: A produce recognition system. In: Proc. 3rd IEEE Workshop on Applications of Computer Vision (WACV 1996), Washington, DC, USA, p. 244. IEEE Computer Society, Los Alamitos (1996)
3. Tao, Y., Morrow, C.T., Heinemann, P.H.: Fourier-based separation technique for shape grading of potatoes using machine vision. Trans. American Society of Agricultural Engineers 38(3), 949–957 (1995)
4. Muir, A.J., Ross, D.W., Dewar, C.J., Kennedy, D.: Defect and disease detection in potato tubers. In: Proceedings of SPIE - The International Society for Optical Engineering, vol. 3543, pp. 199–207 (1999)
5. Tao, Y., Heinemann, P.H., Varghese, Z.: Machine vision for color inspection of potatoes and apples. Trans. American Society of Agricultural Engineers 38(5), 1555–1561 (1995)
6. Heinemann, P.H., Pathare, N.P., Morrow, C.T.: An automated inspection station for machine-cision grading of potatoes. Machine Vision and Applications 9(1), 14–19 (1996)
7. Zhou, L., Chalana, V., Kim, Y.: Pc-based machine vision system for real-time computer-aided potato inspection. International Journal of Imaging Systems and Technology 9(6), 423–433 (1998)
8. Guannan, Z., Yuzhi, T., Junxiong, Z., Wei, L.: Automatic detecting and grading method of potatoes with computer vision. Nongye Jixie Xuebao / Trans. Chinese Society of Agricultural Machinery 40(4), 166–168+1 (2009)
9. Barnes, M., Duckett, T., Cielniak, G., Stroud, G., Harper, G.: Visual detection of blemishes in potatoes using minimalist boosted classifiers. Journal of Food Engineering 98(3), 339–346 (2010)
10. Freund, Y., Schapire, R.: A short introduction to boosting. Journal of Japanese Society for Artificial Intelligence 14(5), 771–780 (1999)
11. De Haan, E.G., van den Bovenkamp, G.W.: Improved diagnosis of powdery scab (spongospora subterranea f.sp. subterranea) symptoms on potato tubers (solanum tuberosum l.). Potato Research 48(1-2), 1–14 (2005)
12. Schapire, R., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. In: Proc. Conf. on Computational Learning Theory, pp. 80–91 (1998)
13. Vezhnevets, A.: GML AdaBoost MATLAB Toolbox (2006)

The Colour Sketch Recognition Interface for Training Systems

Marcin Bernaś

Silesian University of Technology, Department of Transport
ul. Krasińskiego 8, 40-019 Katowice, Poland
marcin.bernas@polsl.pl

Abstract. Presentation, in a form of a sketched graphical diagram, is often used as a training approach of various didactic levels. However, the trainee examination using the hand-sketch diagrams is still not a common practice, due to the evaluation complexity; both for a teacher and for an automatic recognition system. The elaborated interface, supported by the low-level recognition algorithm, simplifies the evaluation procedures. The solution removes both interpretation uncertainty factor and simple errors from the sketch, using the user interaction interface. The users interaction is based on online pattern recognition of strokes in a colour context. Their representation is returned to the user and stored as a graph structure. Additionally, the interface tracks the sketch creation process for more precise final evaluation. The interface can support many fields as various as business, computer science, mathematics, engineering, music and so on; where a set of graphical symbols are lines and curves composition.

1 Introduction

Sketching is a natural way of expressing oneself. It is a simple and fast approach to convey information as a theoretical idea or exact engineer specification. There are many fields, for example UML diagrams design, where sketches are usually made on paper and then transferred to the computer [1]. The transfer is time consuming and can generate additional interpretation errors.

Similar condition can be noticed in the e-learning field. Works sent by students are usually final version of the sketch (not always the best one). Teacher or an automatic sketch recognition system can easily find an error, but not its source. Gathered data usually lack of the creation history, which could be helpful to evaluate the sketch more adequately. Moreover, the open source learning platforms (e.g. Moodle) lack the functional tool for sketch acquisition and pre-processing.

The article proposes a sketch acquisition tool for open source e-learning platforms. The tool is based on Flash environment, due to its multiplatform character, simplicity and popularity among users.

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 217–224, 2010.

2 Model Definition

One of the most difficult issue in creating the sketch acquisition tool and recognition system is to find a balance between recognition effectiveness and drawing freedom. On the one hand if tool restricts drawing to the pre-defined symbols set, the evaluation is simplified at the cost of educational value. On the other hand, if the tool requires learning the specific sketching manner, the user will have to focus on the tool instead of a given task.

The tool offers continuous feedback from the low level recognition system, preparing data for the teacher or high level recognizer in the same time. The tool is giving the user full sketching freedom, presenting in real-time drawing of recognised shapes as a sketch shadow. The user is aware of the pre-evaluation result and can perform corrections if needed.

The proposed solution (fig. 1) concerns: registering sketch creation process as a sequence of pixels, converting the sequence to an elemental stroke graph with a features description, performing pre-recognition process to catch up the simple drawing error and preparing sketch description for export.

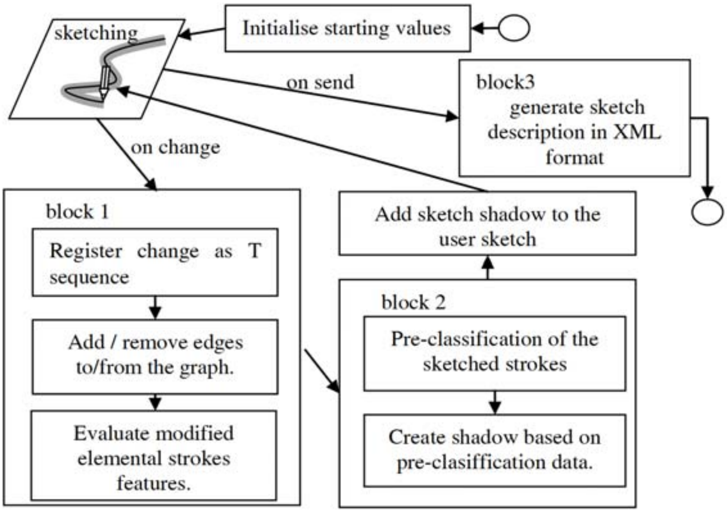

sketching
Initialise starting values
on send
block3
generate sketch description in XML format
on change
block 1
Register change as T sequence
Add / remove edges to/from the graph.
Evaluate modified elemental strokes features.
Add sketch shadow to the user sketch
block 2
Pre-classification of the sketched strokes
Create shadow based on pre-clasiffication data.

Fig. 1. Real-time based sketching tool mechanism

2.1 Block1 - Online Features Gathering

Sketched data are stored in enhanced graph, defined as a triple:

$$G = (V, E, C) \tag{1}$$

where:

V – set of vertices representing pixels in the sketch,

E – set of the edges representing the smallest traceable elemental strokes (e-stroke),

C – features set describing e-strokes.

Each defined pixel (v) of V set is described by both X and Y coordinates, fulfilling following equation:

$$\forall v_i \in V v_i = (x_i, y_i), x_i, y_i \in N \wedge x_i \leq x_{\max} \wedge y_i \leq y_{\max}, \tag{2}$$

where:

x_{max}, y_{max} – width / height of the sketched area in pixels.

E-strokes are the basic units processed by the algorithm. Each e-stroke is described as a pair:

$$\forall e_k \in E e_k = (v_i, v_j) \exists! v_i, v_j \in V \tag{3}$$

Proposed tool tracks every new pixel. Algorithm creates next e-stroke, if Euclidean distance between first unclassified pixel and the currently drawn pixel exceeds s_{min} parameter. Elemental stroke is created within a stroke based on sketched pixels. Each e-stroke is characterised by an equal length. If unclassified pixels mutual distances do not exceed a s_{min} threshold within a stroke, they are not classified.

The stroke ends if a user releases the mouse button and Euclidean distance between last pixel and a new one exceeds s_{min}. Parameter s_{min} reduces the data size at the expense of e-stroke approximation accuracy. The best results were achieved for s_{min} value between 6 and 10. This simplification greatly speeds up extraction and description process presented below.

To create on-line immediate pre-classification tool the number of features should be limited. Concerning works [1], [3] and [10] and features of a commonly used language for high level sketch recognition [2] [6], the following set of features were proposed:

$$C = \{c_{col}, \alpha, \alpha^d, CC, EC\} \tag{4}$$

There are colour (c_{col}), angle (α), relative angle (α^{col}), curvature classification (CC: line, clockwise arc, anti-clockwise arc) and curvature characteristic (EC: characteristic e-stroke and CC feature defining possible begin/end of a stroke) features describing e-stroke. The colour (c_{col}) feature differs the sketch functionality in proposed tool. For example a text, fills or a background can be separated within sketch itself by colours.

Next feature considers application of Rubine's absolute angle between e-stroke and X-axis. Feature is evaluated using the arctangent function for each e-stroke.

Newly proposed feature introduced in this paper is the relative angle of an e-stroke $\alpha^d(e_k)$. This feature is invulnerable to the rotation and can easily be used in robust pre-classification algorithm for simple devices. The relative angle evaluation algorithm is based on considered e-stroke e_k and preceding e-stroke e_l. Preceding e-stroke are selected from stroke sequence (if parameter s_{con}=false), or as the closest e-stroke in drawing sketch, which Euclidean distance between end point of candidate e_l and beginning of a surveyed e_k is smaller than s_{min} parameter. If no candidate can be found the empty e_l edge is sent to the algorithm.

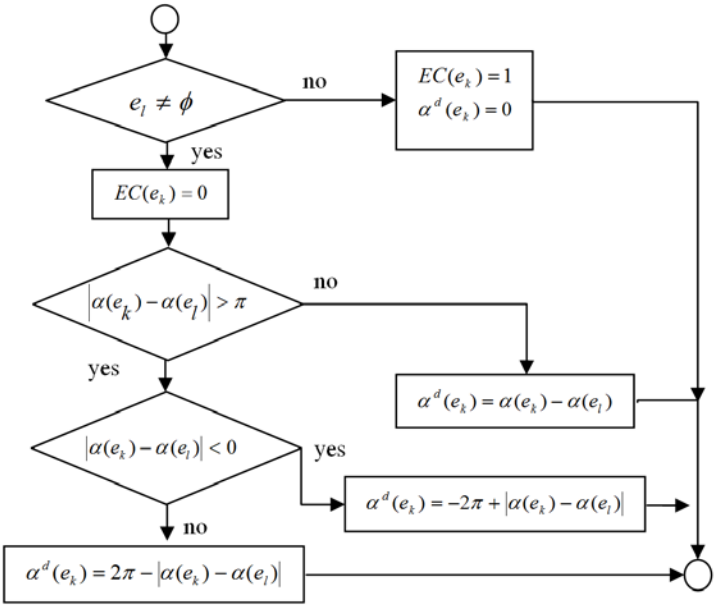

$e_l \neq \phi$
no
$EC(e_k)=1$
$\alpha^d(e_k)=0$
yes
$EC(e_k)=0$
$|\alpha(e_k)-\alpha(e_l)|>\pi$
no
yes
$\alpha^d(e_k)=\alpha(e_k)-\alpha(e_l)$
$|\alpha(e_k)-\alpha(e_l)|<0$
yes
$\alpha^d(e_k)=-2\pi+|\alpha(e_k)-\alpha(e_l)|$
no
$\alpha^d(e_k)=2\pi-|\alpha(e_k)-\alpha(e_l)|$

Fig. 2. Relative angle evaluation algorithm

The $\alpha^d(e_k)$ value is generated using algorithm presented on fig. 2. Additionally the initial values of feature EC are defined.

2.2 Block2 –Pre-evaluation and Beautification Algorithm

The recognition procedure is performed in two stages. The relative angles are quantified initially to the classes of CC feature. Based on the CC feature the final pre-classification of strokes is performed.

Many measures, used by other authors ([4] [9]), could not be applicable, due to the insufficient information where a stroke either begins or ends (immediate recognition). Therefore the pre-classification uses n-size sequence of e-strokes treated as left, right and centre background of considered $e_{(k-1/2n+1)}{=}e_w$ e-stroke (central one). The background for classification is built based on a weights matrices (mk_1 and mk_2):

$$mk_i = [{}^i m_{p,o}] 1 \leq o \leq 3, 1 \leq p \leq n, 1 \leq i \leq 2 \tag{5}$$

CC feature evaluation is delayed by 1/2n to the current e-stroke (e_k). Using matrices and last n e-strokes relative angles (E_k^n) three parameters and its weights for each matrix are evaluated:

$$E_k^n = [\alpha^d(e_{k-n+1}), \alpha^d(e_{k-n+2}), ..., \alpha^d(e_w), ..., \alpha^d(e_{k-1}), \alpha^d(e_k)]^k \tag{6}$$

$$P^i = [{}^k p_1^i, {}^k p_2^i, {}^k p_3^i] = E_k^n.(mk_i)^T \tag{7}$$

where: . – matrix multiplication, T – transpose operation.

$$^{k}PW^{i} = [^{k}pw^{i}_{1}, ^{k}pw^{i}_{2}, ^{k}pw^{i}_{3}], ^{k}pw^{i}_{j} = \sum_{p=1}^{n} {}^{i}m_{p,o}, 1 \leq o \leq 3 \qquad (8)$$

Parameters $^{k}p^{i}_{1}$ and $^{k}p^{i}_{3}$ are respectively left-side and right-side background for a given e-stroke. Parameter $^{k}p^{i}_{2}$ value, which represents the e-stroke itself (e_w), allows removing small errors and noise created while dividing stroke into smaller parts. The mk_i weights are selected based on both the effectiveness analysis and a guidelines stated in [7] [8] [9]. Weights matrix are used for curvature classification (mk_1) as well as for new strokes starts/ends finding (mk_2). The matrices values are shown in fig. 3. The mk_1 asymmetry decreases noise factor, while evaluating CC feature value in proximity to detected characteristic point. The mk_2 usually holds a gap between considered e-stroke and its left / right neighbourhood. A survey shows [9] that edges near characteristic points contain the biggest noise value. In case of mk_2 matrix, the considered edge is dominant one. Only strong neighbourhood values can influence change to the classification.

Both CC and EC feature evaluation algorithms are tuned using two parameters: er_1, defining maximal curvature of straight line, and er_2, defining sensitivity of characteristic point extraction.

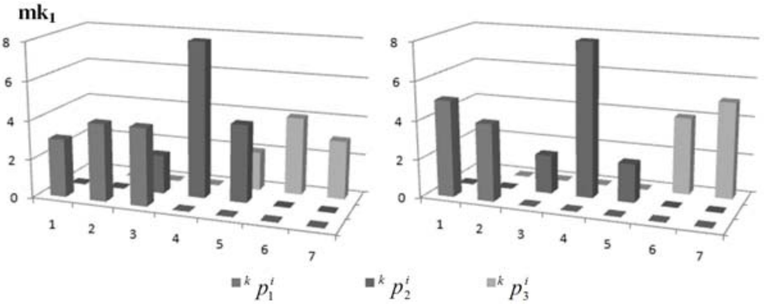


Fig. 3. Matrices of weight - graphical representation

If $EC(e_w)$ feature value is not 1 (beginning of a stroke), the standard curvature type is evaluated. The CC feature classification algorithm of a curvature is described by a following equation:

$$trends = \sum_{i=l}^{3} \frac{^{k}p^{1}_{i}}{^{k}pw^{1}_{i}}, where: l = \begin{cases} 1: \forall e_u, u \in (k-n+1...w) EC(e_u) = 0 \\ 2: \exists e_u, u \in (k-n+1...w) EC(e_u) > 0 \end{cases} \qquad (9)$$

$$CC(e_w) = \begin{cases} 2: trends \leq -er_1 \\ 3: \ \ trends \geq er_1 \\ 4: |trends| < er_1 \end{cases} \qquad (10)$$

The prediction of curvature of e-strokes, between e_{w+1} and e_k, is evaluated using both equation (10) and ${}^kp_3^1/{}^kpw_3^1$ as trends parameter.

Parameters ${}^kP^2$ and ${}^kPW^2$ are used for classification of EC feature value. The EC feature is evaluated based on CC values, therefore the EC feature evaluation, in the same loop, is delayed by 1 (e_{w-1}). Characteristic points are represented by a positive values of EC feature and they are defined by the following tests, together with its wages, presented in table 1. If more than one test is positive, the class with the highest weight is chosen (1-the highest weight). At this stage, EC feature can obtain value between 0 and 8 representing characteristic points types.

In case of strong curve deviation, one characteristic point can be registered in more than one e-stroke. To focus it in one e-stroke the following deviation parameter is measured:

$$dev(e_w) = \left| \frac{{}^kp_2^2}{{}^kpw_2^2} - \frac{{}^kp_1^2}{{}^kpw_1^2} \right| + \left| \frac{{}^kp_2^2}{{}^kpw_2^2} - \frac{{}^kp_3^2}{{}^kpw_3^2} \right| \tag{11}$$

If a characteristic point of the same curvature weight (defined in table 1) was detected in both e_w and preceding $\{e_{w-1}, e_{w-2}, \ldots, e_{w-n/2+1}\}$, the characteristic point with smaller deviation value (dev) is changed to a regular stroke.

The beautification algorithms can be turn on/off depending on the device performance and available memory. To draw arc as a feedback the Bezier curves were used. For arcs being part of circle (total curvature exceed $\pi/2$), the algorithm from [9] were used. Lines are drawn using the linear approximation.

2.3 Block3 –XML Data Generation Format

To export sketch in a format accepted by higher level recognition systems e.g. [2] or [6] all description features were implemented: line, arc, parallel, length, connected, meets as well as orientation feature. Elemental shapes are separated from each other by characteristic points defined in table 1. The description based on the multiple language differs, however using XSLT templates it can be easily converted.

3 Experiments and Summary

The presented tool is designed for devices supporting the browser technology with Flash Player 8 plug-in or higher. The application is parameterised by the manifest, stored in the current directory on the server. User can change default values as e-stroke size (s_{min}) and basic tuning parameters (er_1 and er_2) using a graphical tutor. Manifest consists up to 20 colour definition along with its function in the sketch as well as the task description. The manifest also contains the address of a server to receive data along with the user identification number. Data is send and stored as a XML file. The file is accessed by a teacher using the same tool. Added changes can be returned to the user as a feedback.

Table 1. Characteristic points test

Test (true/false)	Comment / wt.
$curv = \frac{\sum_{u=k}^{k-n} \frac{{}^{u}p_2^2}{{}^{u}pw_2^2}}{n}$	average curvature of preceding e-strokes / -
$er_{\max} = er_2 * curv$	characteristic point threshold / -
$tmp : \left\lvert \frac{\frac{{}^{k-1}p_2^2}{{}^{k-1}pw_2^2}}{curv} \right\rvert < \frac{1}{er_2} \vee \left\lvert \frac{\frac{{}^{k-1}p_2^2}{{}^{k-1}pw_2^2}}{curv} \right\rvert > er_2$	- / -
$t_1 : CC(e_{w-1}) = 2 \wedge CC(e_{w-2}) = 3 \wedge \frac{{}^{k-1}p_1^2}{{}^{k-1}pw_1^2} > 0 \wedge \frac{{}^{k-1}p_3^2}{{}^{k-1}pw_3^2} < 0$	curvature change from anti-clockwise to clockwise / 2
$t_2 : CC(e_{w-1}) = 3 \wedge CC(e_{w-2}) = 2 \wedge \frac{{}^{k-1}p_1^2}{{}^{k-1}pw_1^2} < 0 \wedge \frac{{}^{k-1}p_3^2}{{}^{k-1}pw_3^2} > 0$	curvature change from clockwise to anti-clockwise / 3
$t_3 : \begin{array}{c} tmp \wedge CC(e_{w-1}) = 4 \wedge (CC(e_{w-2}) = 2 \vee CC(e_{w-2}) = 3) \wedge \\ \wedge \left\lvert \frac{{}^{k-1}p_3^2}{{}^{k-1}pw_3^2} \right\rvert * er_2 < \left\lvert \frac{{}^{k-1}p_1^2}{{}^{k-1}pw_1^2} \right\rvert \end{array}$	curvature change from curve to a line. / 2
$t_4 : \begin{array}{c} tmp \wedge (CC(e_{w-1}) = 2 \vee CC(e_{w-1}) = 3) \wedge CC(e_{w-2}) = 4 \wedge \\ \wedge \left\lvert \frac{{}^{k-1}p_1^2}{{}^{k-1}pw_1^2} \right\rvert * er_2 < \left\lvert \frac{{}^{k-1}p_3^2}{{}^{k-1}pw_3^2} \right\rvert \end{array}$	curvature change from line to a curve. / 2
$t_5 : \begin{array}{c} CC(e_{w-2}) = 2 \wedge (CC(e_{w-1}) = 2 \vee CC(e_w) = 2) \wedge \\ \wedge \frac{{}^{k-1}p_3^2}{{}^{k-1}pw_3^2} < -er_{\max} \end{array}$	strong deviation in a clockwise curve / 2
$t_6 : \begin{array}{c} CC(e_{w-2}) = 3 \wedge (CC(e_{w-1}) = 3 \vee CC(e_w) = 3) \wedge \\ \wedge \frac{{}^{k-1}p_3^2}{{}^{k-1}pw_3^2} > er_{\max} \end{array}$	strong deviation in a anti-clockwise curve /2
$t_7 : tmp \wedge \left\lvert \frac{{}^{k-1}p_2^2}{{}^{k-1}pw_2^2} \right\rvert > er_{\max}$	very strong deviation / 1

The XML data can be also automatically processed by high level recognition algorithms, e.g. LADDER[6]. Performed experiments considered the circuit creation, text matching and tunes description. The 86% recognition accuracy, for 200 basic shape samples, was achieved for er_1=0.04 and er_2=2.5 parameters. Generated information was sufficient for high-level automatic recognitions with 96% average detection rate. A lower detection rate in comparison to the other solutions [3] [4] [5] [6] is a cost of the single and multi-stroke recognition as well as the multiple shapes recognition within one stroke. The adaptation of the solution to one type of sketch would improve the efficiency significantly. Next step is to create the high level recognition tool, which could benefit from all *EC* classes types defined in described tool.

Described tool is a open source project and we encourage the researchers to use and contribute to its development. Tool is available at http://www.science-tech.pl/e-tools.

References

1. Rubine, D.: Specifying Gestures by Example. Computer Graphics 25(4) (July 1991)
2. Alvarado. C., Oltmans, M., Davis, R.: A Framework for Multi-Domain Recognition. AAAI Technical Report SS-02-08 (2002)
3. Blagojevic, R., Schmieder, P., Plimmer, B.: Towards a Toolkit for the Development and Evaluation of Sketch Recognition Techniques. In: IUI 2009 Workshop on Sketch Recognition, Florida (2009)
4. Li, J., Zhang, X., Ao, X., Dai, G.: Sketch Recognition with Continious Feedback Based On Incremental Intention Extraction. In: IUI 2005 Workshop on Sketch Recognition, San Diego, California, USA (2005)
5. Plimmer, B., Freeman, I.: A Toolkit Approach to Sketched Diagram Recognition. In: People and Computers XXI – HCI, Proceedings of HCI (2007)
6. Hammond, T., Davis, R.: LADDER: A Sketch Recognition Language. MIT Computer Science and Artificial Intelligence Laboratory (September 2004)
7. Mahoney, J., Frommerz, M.: Three main concerns in sketch recognition and an approach to addressing them. In: Sketch Understanding. AAAI Spring Symposium, Stanford, California, pp. 105–112. AAAI Press, Menlo Park (2002)
8. Hammond, T.: Enabling Instructors to Develop Sketch Recognition Applications for the Classroom. In: 37th ASEE/IEEE Frontiers in Education Conference. WI (2007)
9. Paulson, B., Hammond, T.: PaleoSketch: Accurate Primitive Sketch Recognition and Beatification. In: IUI 2008, Maspalomas, Gran Canaria, Spain (2008)
10. Liao, S., Wang, X., Lu, J.: An Incremental Bayesian Approch To Sketch Recognition. In: Proc. of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou (2005)

Interactive Hydraulic Erosion Using CUDA

Richard Bezin[1], Alexandre Peyrat[1], Benoit Crespin[1], Olivier Terraz[1],
Xavier Skapin[2], and Philippe Meseure[2]

[1] XLIM - UMR 6172 - CNRS, University of Limoges, France
[2] XLIM-SIC, University of Poitiers, France
benoit.crespin@xlim.fr

Abstract. This paper presents a method to simulate hydraulic erosion and sedimentation on a terrain represented by a triangular mesh in real-time. Our method achieves interactive performances by dynamically displacing vertices using CUDA following physically-inspired principles; the mesh is generated in a preprocessing step to avoid degenerated cases in highly deformed areas.

1 Introduction

Fluvial processes study how landforms are created by rivers and streams through erosion, sediment transport and deposit over time. In this paper we're interested in reproducing the hydraulic erosion process due to water, neglecting chemical dissolution and other marginal processes. Our goal is to automatically obtain visually realistic terrains eroded by water flows, for example by deepening a valley due to *stream erosion*. Since hydraulic erosion is mainly due to maximal flood levels rather than normal activity [1], dynamic control of the rate of flow is desirable. The main contribution of our approach is its ability to produce plausible results at an interactive framerate even for very large scenes, by implementing our method within the CUDA framework [2]. As a consequence, some choices are inherent to this implementation, such as using a particle system to represent the water flow, which performs well with highly parallelized architectures, although real-time fluid simulations may rely on other discretization models. A snapshot of our application is shown on Fig. 3.

Another consequence is that some processes may be neglected or simplified in order to maintain interactive rates, if their visual contribution does not appear significant in the final result. As stated earlier, chemical dissolution for instance does not have a significant impact on the visually perceptible details due to hydraulic erosion [1]. Other processes such as sediments acting erosively on the surface would be too computationally expensive for any existing method if we were to represent each rolling and sliding grain in the flow.

This paper is organized as follows. We first describe recent works in the literature addressing the problem of hydraulic erosion in section 2. Section 3 describes our hydraulic erosion model and its implementation with CUDA; an adaptive heightfield generation method is also presented to reduce the amount of triangles in a preprocessing step. Finally, results obtained with our application are shown in section 4.

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 225–232, 2010.

2 Previous Work

Recent works focus on terrain *modifications* by hydraulic processes, by explicitly representing fluid motion and interactions between the fluid and the initial terrain model. Representations based on voxels [3,4] divide the terrain into a set of small 3D cubes. Each voxel stores some information about the amount of material it contains, its geological resistance to hydraulic erosion, etc. In [5], visually plausible concave surfaces are created through erosion using discrete surface curvature obtained from the voxel grid. Navier-Stokes equations can be solved by different numerical methods. A semi-Lagrangian method is used in [6] to run on simple scenes in real-time. The capacity C for a particle to transport sediment depends on its velocity: if C is above a predefined threshold, it can erode the terrain by sweeping sediment away, otherwise it may deposit sediment. Other recent approaches take into account the local slope of the terrain and evaporation [7], or different types of material [8]. Smoothed Particle Hydrodynamics (SPH) are used in [9] to represent fluid motion, with the terrain modelled as a triangulated heightfield. Interactions between the fluid and the terrain are computed through the use of static particles sampling the triangulated mesh, which exchange data with SPH particles: erosion is achieved by transferring material from static particles to SPH particles, whereas sediment deposition is the inverse process, and terrain modifications are then obtained by vertical extrusion of the corresponding triangles. Simulations conducted with this approach run in interactive time with up to 25K particles thanks to a GPU implementation.

However, even if results obtained with these methods are visually spectacular, their high computation and memory costs make them either prohibitive for real-time applications or limited to simple scenes. Therefore our goal is to get the same quality with more approximations. A good example is the fluid discretization model: SPH is a well-known method for accurately simulating small-scale details, but other particle-based fluid models may provide better results to enforce incompressibility [10]. By relying on a similar philosophy, we aim at approximating erosion and fluid-terrain exchanges in order to generate a visually realistic but not necessarily precise simulation.

3 Hydraulic Erosion Model

We use a particle-based fluid model, relying on the n-body gravitational simulation provided within CUDA where particle-particle interactions are computed with a DEM method [2]. We also choose to represent the terrain as a non-regular triangular mesh; triangles are obviously the best choice for real-time simulations, but implies some kind of level-of-details process to avoid degenerated triangles in strongly eroded areas, as presented in section 3.3. The main loop of our simulation method can be summarized as:

1. Update fluid particles positions
2. Calculate collisions and accumulate erosion on triangles

3. Calculate sediment deposition from fluid particles and accumulate sedimentation on triangles
4. Update vertices' positions and remove unnecessary particles

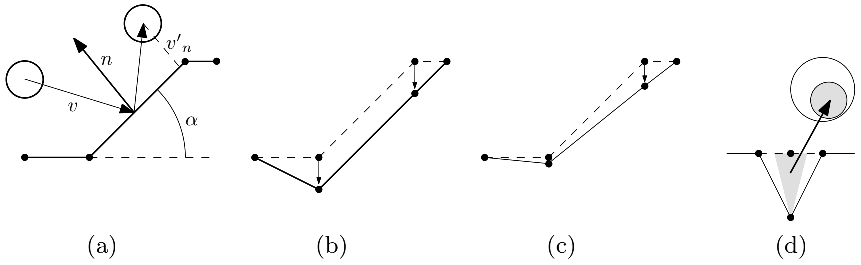
We describe in the following how to compute collisions and update vertices based on erosion and sedimentation processes.

3.1 Particle/Triangle Collision

In order to obtain a realistic fluid flow, we need to detect and handle collisions between particles and triangles. Our method inspired by [11] computes segment/triangle intersections. In our case this segment is given by the successive positions of a fluid particle p_t and p_{t+1}, and we keep the closest triangle to p_t. Velocity $\boldsymbol{v}$ is decomposed in $\boldsymbol{v_t}$ and $\boldsymbol{v_n}$, its tangential and normal velocity relative to the normal vector $\boldsymbol{n}$ respectively; after collision, normal velocity $\boldsymbol{v'_n}$ is reduced to simulate damping and force particles to slip over the surface as shown on Fig. 1a. When colliding with the triangle, the particle erodes it by a certain amount C (as in [7]):

$$C = K_c \cdot \sin\alpha \cdot |\boldsymbol{v}| \tag{1}$$

where K_c is the erosion rate and α is the triangle's tilt angle, meaning that horizontal areas are less eroded. To apply erosion we then need to displace its vertices downwards: if vertices are uniformly displaced by the same amount $C/3$, the slope of the triangle is preserved but unrealistic results may appear (see Fig. 1b). We choose to apply a larger weight to the highest vertex, which produces more gentle slopes as shown on Fig. 1c. The final displacement applied to a vertex is the sum of displacements computed from each triangle to which it belongs.



(a) (b) (c) (d)

Fig. 1. (a) Collision handling between a particle and a segment in 2D **(b)** Uniform vertical displacement of endpoints **(c)** Weighted displacement depending on endpoints' height values **(d)** 2D representation of the local volume eroded from one vertex (in gray) and transferred to the nearest fluid particle

3.2 Erosion and Sedimentation

Displacements due to hydraulic erosion are applied to update vertices positions as described above, which in turn generate sediments that will be transported by the flow. Actually displacements are not taken into account at each simulation step, instead we choose to accumulate displacements during n frames before applying erosion every n steps only. This implies that triangle-particle collisions are not as accurate as possible, however re-hashing the entire mesh at each step would be too expensive even on GPU. Another problem when applying erosion at each step would be that only small-sized sediment particles are generated, although sediments are usually composed of mineral rocks or sands of various sizes, depending on the composition of the river bed [1]. Parameter n, as well as all the other parameters of our model, can be modified during the simulation.

We derive the size of a sediment eroded from the bed at a given vertex from the approximated volume of the cone that is "lost" due to erosion as shown on Fig. 1d. This spherical sediment of size s is then transported by the flow, and will eventually be deposited back to the bed. Implementing this transport/deposition process is complicated, since it depends on many parameters, including sediment size and shape, mineral composition and flow velocity. The well-known Hjulström curve for example [1] rules that, for the same flow velocity, small-sized grains are transported for a longer period of time. Because computing exact flow velocity is a time-consuming task, we simplify the problem by considering the *expected transport time* ε of a sediment:

$$\varepsilon = \lceil \varepsilon_M (1 - \frac{s^x}{s_M^x}) \rceil \tag{2}$$

where ε_M and s_M represent the maximal transport time and the maximal size of a sediment particle respectively, and $x = 1/s_M$.

Newly created sediment grains are advected with the flow by attaching them to the nearest fluid particle. At each subsequent step of the simulation their transport time ε is decremented: when it reaches zero the sediment falls from its fluid particle down to the river bed. Eq. 2 ensures that near-zero size sediments will be transported during ε_M steps before being deposited, whereas large sediments of size s_M will only have one step in a flow with the same velocity. If the fluid particle slows down, the sediment will be transported on a shorter distance, which is consistent with Hjulström's rule.

Sedimentation means transferring the volume of the sediment to the nearest triangle below its attached fluid particle. As with erosion, this volume is distributed among the three vertices but those are displaced upwards such that the gained volume is equivalent, using different weights (a similar rule can be found in [9]). Fig. 3 shows the application of our algorithms with eroded or sedimented triangles colored in purple. The iterative nature of our model is visible on Fig. 4.

3.3 Implementation with CUDA and Adaptive Heightfield Generation

Our implementation runs entirely on GPU and can generate particles through emitters or deactivate particles if they evaporate or flow out of the simulation. Particles are stored in a hashing grid to accelerate neighborhood queries. An auxiliary hashing grid storing triangles is used for efficient particle-triangle collision detection [12], and is updated every n steps when erosion or sedimentation is applied (*ie* when vertices are displaced).

The goal with GPU programming is to reduce concurrent access to read or write data. If we consider that a particle may erode or sediment only one triangle at each step, our main problem is to efficiently compute which particles contribute to the displacement applied to each vertex. We give details about this computation below, vertices and particles displacements described in sections 3.1 and 3.2 being easier to implement. To handle particle-triangle interactions, we extend the work presented in [13] based on four successive steps implemented through atomic operations to ensure maximal performances; Table 1 describes an example where particles respectively indexed 0 and 2 collide with triangles defined by vertices $(4, 6, 8)$ and $(2, 6, 8)$. Collisions are stored in two arrays indexParticle and indexVertex (step **a**), then these arrays are sorted regarding indexVertex (step **b**). Two arrays cellStart and cellEnd are filled using the first and last index of each vertex stored in indexVertex (step **c**); for example, vertex 6 starts at index 2 and stops at index 4 in indexVertex. Finally particles which contribute to displace each vertex v are obtained by considering the sorted array indexParticle from cellStart[v] to cellEnd[v]-1 (step **d**); for vertex 6 these are particles 0 and 2, stored in indexParticle at indices 2 and 3.

Table 1. Finding particles that contribute to the displacement applied to each vertex

	0	1	2	3	4	5	6	7	8	...
indexParticle	0	0	0	1	1	1	2	2	2	...
indexVertex	4	6	8				2	6	8	...

(a)

	0	1	2	3	4	5	6	7	8	...
indexVertex	2	4	6	6	8	8				...
indexParticle	2	0	0	2	0	2				...

(b)

	0	1	2	3	4	5	6	7	8	...
cellStart			0		1		2		4	...
cellEnd			1		2		4		6	...

(c)

vertex	0	1	2	3	4	5	6	7	8	...
particles			2		0		0,2		0,2	...

(d)

To get realistic results, we need to refine triangles in strongly eroded areas to avoid visually perceptible artifacts due to initial mesh resolution. This well-known problem [14] implies recursive subdivisions of neighbouring triangles, but we can't refine the mesh on the fly without degrading performances or increasing memory consumption to store neighbouring relationships. We choose to implement an offline pre-processing pass to generate more details where the erosion is likely to occur. Generating larger triangles in areas where less erosion is expected is also a way to increase performances since there will be less triangles to handle in the simulation. Fortunately, detecting areas where erosion may be strong is relatively easy because the fluid always flows downwards; selection criteria for a triangle to be subdivided are thus its depth and its gradient vector given by the heightfield using vertices interpolation. Therefore the highest gradient and depth in a region, the more details we add (see Fig. 2).

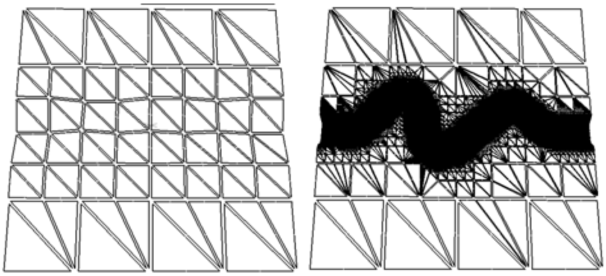


Fig. 2. Different levels of mesh subdivision

4 Results and Conclusion

After the terrain is subdivided by the method above and the user has set particles emitters, simulation starts and lets the user modify all parameters in real-time through sliders (time step, erosion rate, sediment capacity constant, etc.), as shown on Fig. 3 (top-left). The snapshot presented on Fig. 3 shows our interactive application with multiple fast-flowing streams with 130k particles running down a mountain terrain with 80k triangles at approx. 20 fps on a NIVIDIA Quadro FX 3700M. As a comparison, 1.38 seconds per frame are reported in [9] for the same number of particles, but these timings include a basic rendering of the fluid's surface. Another example is shown on Fig. 4 where the initial mesh presented on Fig. 2 (right) is deepened by a running river, and rendered offline on a supercomputer with 3x Tesla C1060 processors at approx. 9 seconds per frame for 20k triangles and 130k particles.

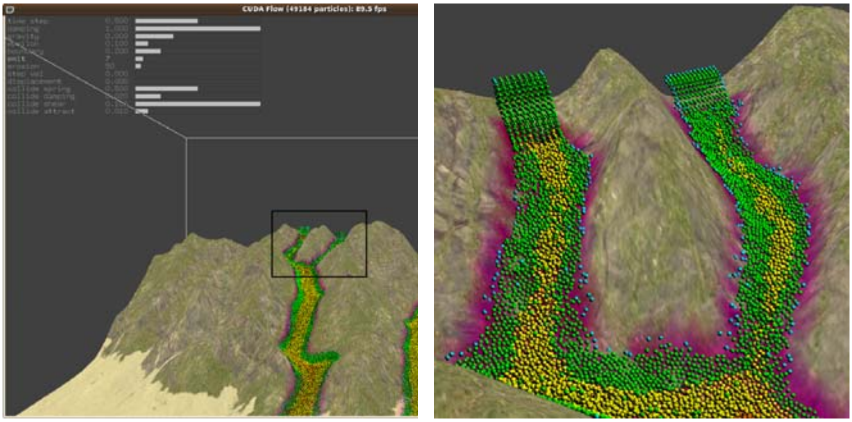


Fig. 3. Left. Streams falling down and eroding a mountain in real-time. **Right.** Zoom on a specific area, showing affected triangles in purple and particles coloration given by their position in the flow.

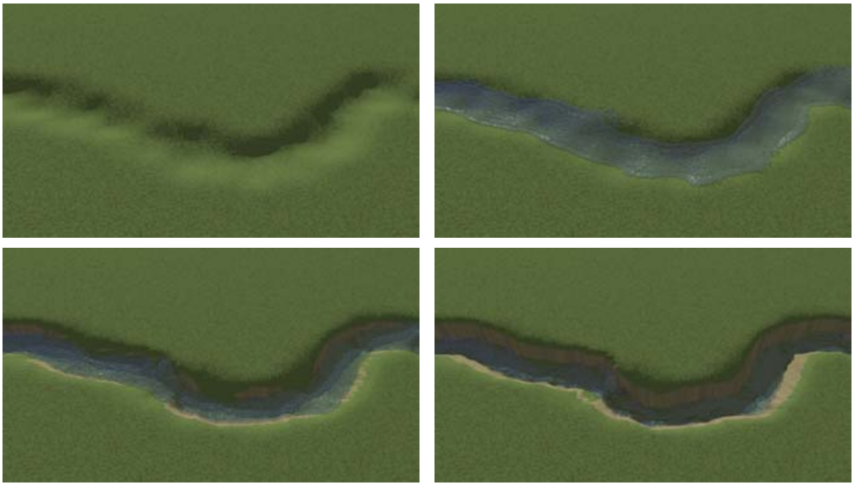


Fig. 4. Canyon progressively deepened by a river

Our method can be considered as physically-inspired, since it relies on physical principles such as gravity and friction but neglects other phenomena which would be too computationally expensive. Results are visually realistic and obtained in real-time; we are able to increase the number of particles in the simulation significantly compared to previous approaches. Integrating other phenomena related to erosion in the future should be relatively straightforward, for instance rain or sediment diffusion into the fluid (as in [9]). Handling dynamic topological changes occurring with hydraulic erosion (some parts of a cliff falling into a river is a good example) will require more complex mesh representations such as a 3D topological model, and an extra amount of work to maintain interactive rates. Another example is *lateral erosion*, which has an impact for example for valleys deepened by rivers. Unlike most existing methods which only consider vertical erosion, in our approach we can choose to displace vertices along their normal vector instead of the vertical axis. However lateral erosion has to be limited because it can involve complex topological problems, such as merging or splitting material volumes.

References

1. Hiscock, K.: Hydrogeology: Principles and Practice. Wiley-Blackwell, Chichester (2005)
2. Nguyen, H. (ed.): GPU Gems 3. NVIDIA Corporation (2008)
3. Dorsey, J., Edelman, A., Jensen, H.W., Legakis, J., Pedersen, H.K.: Modeling and rendering of weathered stone. In: SIGGRAPH, pp. 225–234 (1999)
4. Ozawa, N., Fujishiro, I.: A morphological approach to volume synthesis of weathered stone. In: Volume Graphics, pp. 367–378 (1999)

5. Jones Michael, D., McKay, F., Joseph, B., Matthew, B.: Directable weathering of concave rock using curvature estimation. IEEE Transactions on Visualization and Computer Graphics, 81–94 (2009)
6. Neidhold, B., Wacker, M., Deussen, O.: Interactive physically based fluid and erosion simulation. In: Eurographics Workshop on Natural Phenomena (2005)
7. Mei, X., Decaudin, P., Hu, B.-G.: Fast hydraulic erosion simulation and visualization on gpu. In: Pacific Graphics, pp. 47–56 (2007)
8. Št'ava, O., Beneš, B., Brisbin, M., Křivánek, J.: Interactive terrain modeling using hydraulic erosion. In: Symposium on Computer Animation, pp. 201–210 (2008)
9. Kristof, P., Beneš, B., Krivanek, J., Ondrej, S.: Hydraulic erosion using smoothed particle hydrodynamics. In: Eurographics, pp. 219–228 (2009)
10. Sin, F., Bargteil, A.W., Hodgins, J.K.: A point-based method for animating incompressible flow. In: Symposium on Computer Animation, pp. 247–255 (2009)
11. Badouel, D.: An efficient ray-polygon intersection, Graphics Gems, Vol. I, pp. 390–393 (1990)
12. Kalojanov, J., Slusallek, P.: A parallel algorithm for construction of uniform grids. In: High Performance Graphics, pp. 23–28 (2009)
13. Green, S.: Particle-based fluid simulation for games. In: Game Developers Conference (2008)
14. Von Herzen, B., Barr, A.H.: Accurate triangulations of deformed, intersecting surfaces. In: SIGGRAPH, pp. 103–110 (1987)

Smoothing, Enhancing Filters in Terms of Backward Stochastic Differential Equations

Dariusz Borkowski

Faculty of Mathematics and Computer Science, Nicolaus Copernicus University
ul. Chopina 12/18, 87-100 Toruń, Poland
dbor@mat.uni.torun.pl

Abstract. In this paper we propose a novel approach for reconstruction of images with additive Gaussian noise. In order to solve this inverse problem we use backward stochastic differential equations. Model of the image reconstruction is driven by two stochastic processes. One process has values in domain of the image, and second one in codomain. Appropriate construction of these processes leads to smoothing (anisotropic diffusion) and enhancing filters. Our numerical experiments show that the new algorithm gives very good results and compares favourably with classical Perona-Malik method.

1 Introduction

Let D be a bounded, convex domain in $\mathbf{R}^2$, $u : \overline{D} \to \mathbf{R}$ be an original image[1] and $u_0 : \overline{D} \to \mathbf{R}$ be the observed image of the form $u_0 = u + \eta$, where η stands for a white Gaussian noise. We assume that u and u_0 are appropriately regular. We are given u_0, the problem is to reconstruct u. This is a typical example of an inverse problem [2].

The inverse problem of restoration of noisy image by automatic and reliable methods belongs to the most intensively studied topics of image processing. Various techniques of noise removal were proposed to tackle this problem. One may quote the linear filtering, wavelets theory, variational/PDE-based approaches and stochastic modelling [3].

Stochastic methods are generally based on the Markov field theory. Only a few papers involve advanced tools of stochastic analysis such as stochastic differential equations [4,10,11]. Backward stochastic differential equations (in short BSDEs) represent a new direction in image processing. In the literature we can find only theoretical results of BSDE-based applications to image processing [1].

In this paper we explore practical aspects of models based on BSDEs. Appropriate construction of these equations leads to smoothing and enhancing filters, which can be successfully used in image reconstruction. Moreover, we show that the obtained results are favourably with deterministic PDE methods.

[1] In practice we consider images as functions defined on rectangle with values in $[0, 255]$.

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 233–240, 2010.

2 Mathematical Preliminaries

2.1 Skorokhod Problem

Let $D \subset \mathbf{R}^n$ be a domain with closure $\overline{D}$ and boundary ∂D and let y be a function with values in $\mathbf{R}^n$, $y_0 \in \overline{D}$. It is often the case that one wishes to construct a function x with the following heuristic description. The starting point of y and x are the same, so $x_0 = y_0$. If y_t is in $\overline{D}$, then the evolution of x mimics y. If $y \in \partial D$ and imitating y would force x to exit $\overline{D}$ then the correction term is added. This term is the minimal push needed to keep x in $\overline{D}$. If imitating y does not result in the exit of x from $\overline{D}$, then correction term is unnecessary.

Constraint mechanisms of this type occur in statistics, economics, queueing theory, telecommunication and computer networks. The Skorokhod problem provides a very useful definition which translates the heuristic description just given into precise mathematical term. The definition of the Skorokhod problem can be made general enough to cover a very wide variety of domains, constraints and input functions. When the mapping is well defined and sufficiently regular, it provides a convenient tool for the study of many types of constrained deterministic and stochastic problems.

Let $T > 0$ and by $\mathbf{C}([0,T];\mathbf{R}^n)$ we denote the set of $\mathbf{R}^n$ valued continuous functions $f : [0,T] \to \mathbf{R}^n$.

Definition 1. *Let $y \in \mathbf{C}([0,T];\mathbf{R}^n)$, $y_0 \in \overline{D}$. A pair $(x,k) \in \mathbf{C}([0,T];\mathbf{R}^{2n})$ is called a solution to the Skorokhod problem associated with y and D if*

1. *$x_t = y_t + k_t, \quad t \in [0,T]$,*
2. *$x_t \in \overline{D}, \quad t \in [0,T]$,*
3. *k is a function with bounded variation $|k|$ on $[0,T]$, $k_0 = 0$ and*

$$k_t = \int_0^t n_s \, d|k|_s, \ |k|_t = \int_0^t 1_{\{x_s \in \partial D\}} \, d|k|_s, t \in [0,T],$$

where $n_s = n(x_s)$ is an inward normal unit vector at $x_s \in \partial D$.

It is known that if D is a convex set, then there exists a unique solution to the Skorokhod problem [9].

2.2 Fundamentals of Stochastic Analysis

Definition 2. *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space.*

1. *An n-dimensional stochastic process $X = \{X_t; t \in [0,T]\}$ is a parametrised collection of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with values in $\mathbf{R}^n$.*

   For each fixed $\omega \in \Omega$ the function $t \to X_t(\omega)$ is called a trajectory of X and is denoted by $X(\omega)$.

2. *A filtration $(\mathcal{F}_t) = \{\mathcal{F}_t; t \in [0,T]\}$ is a nondecreasing family of sub-σ-fields of $\mathcal{F}$ i.e. $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $0 \le s < t \le T$.*
   By $(\mathcal{F}_t^X)$ we denote a filtration generated by process X, i.e. $\mathcal{F}_t^X = \sigma(X_s; 0 \le s \le t)$.
3. *A stochastic process X is adapted to the filtration $(\mathcal{F}_t)$ (X is $(\mathcal{F}_t)$ adapted) if for each $t \in [0,T]$, X_t is $\mathcal{F}_t$ - measurable random variable.*

Definition 3. *Let Y be $(\mathcal{F}_t)$ adapted process with continuous trajectories, $Y_0 \in \overline{D}$. We say that a pair (X,K) of $(\mathcal{F}_t)$ adapted processes is a solution to the Skorokhod problem associated with Y and D, if for almost every $\omega \in \Omega$, $(X(\omega), K(\omega))$ is a solution to the Skorokhod problem associated with $Y(\omega)$ and D.*

In what follows, by $W = \{W_t; t \in [0,T]\}$ we will denote an n-dimensional Wiener process starting from zero. We assume that we are given a point $x_0 \in \overline{D}$ and a function $\sigma : \mathbf{R}^n \to \mathbf{R}^n \times \mathbf{R}^n$.

Definition 4. *Let Y be an $(\mathcal{F}_t)$ adapted process. A pair $(X, K^{\overline{D}})$ of $(\mathcal{F}_t)$ adapted processes is called a solution to reflected stochastic differential equation SDE (in short reflected SDE)*

$$X_t = x_0 + \int_0^t \sigma(X_s)\, dW_s + K_t^{\overline{D}}, \ \ t \in [0,T], \tag{1}$$

if $(X, K^{\overline{D}})$ is a solution to the Skorokhod problem associated with

$$Y_t = x_0 + \int_0^t \sigma(X_s)\, dW_s, t \in [0,T] \text{ and } D.$$

The process X satisfying (1) we will call the process with reflection. The proof of existence and uniqueness of the solution to reflected SDEs can be found in [9].

Consider the following numerical scheme

$$X_0^m = X_0,\ X_{t_k}^m = \Pi_{\overline{D}}[X_{t_{k-1}}^m + \sigma(X_{t_{k-1}}^m)(W_{t_k} - W_{t_{k-1}})],\ k = 1,2,...,m, \tag{2}$$

where $t_k = \frac{kT}{m}$, $k = 0,1,...,m$ and $\Pi_{\overline{D}}(x)$ denotes a projection of x on the set $\overline{D}$. Since D is convex, the projection is unique.

Theorem 1. *Let $(X, K^{\overline{D}})$ be the solution to the reflected SDE (1). If there exists $C > 0$ such that $\|\sigma(x) - \sigma(y)\|^2 \le C|x-y|^2$, then*

$$\lim_{m \to +\infty} |X_T^m - X_T| = 0 \quad \mathcal{P} - a.s.$$

The proof of the above theorem can be found in [8].

Let $(\mathcal{F}_t^W)$ be a filtration generated by a l-dimensional Wiener process W, $\xi \in \mathbf{L}^2(\Omega, \mathcal{F}_T, P, \mathbf{R}^k)$ and $f : \Omega \times [0,T] \times \mathbf{R}^k$ be a Lipschitz continuous in the space variable.

Definition 5. *The solution to the BSDE associated with ξ anf f is a pair of $(\mathcal{F}_t^W)$ - measurable processes (Y_t, Z_t) with values in $\mathbf{R}^k \times \mathbf{R}^{k\times l}$ satisfying the following conditions:*

$$\mathbf{E}\left[\int_0^T \|Z_s\|^2\, ds\right] < \infty,$$

$$Y_t = \xi + \int_t^T f(s, Y_s) ds - \int_t^T Z_s\, dW_s, \quad t \in [0, T]. \tag{3}$$

See [6] for the proof of existence and uniqueness of the solution to BSDEs.

Let us now consider the following numerical scheme to the BSDE (3):

$$Y_{t_m}^m = \xi^m, \quad \hat{Y}_{t_k}^m = \mathbf{E}[Y_{t_{k+1}}^m | \mathcal{F}_{t_k}^m], \quad Y_{t_k}^m = \hat{Y}_{t_k}^m + \frac{T}{m} f(t_k, \hat{Y}_{t_k}^m), \tag{4}$$

where $t_k = \frac{k\,T}{m}$, $k = 0, 1, ..., m-1$, $\mathbf{E}$ denotes the expected value and $\mathcal{F}^m$ is the filtration generated by discretization of the Wiener process.

Theorem 2. $\sup_{0\leq k\leq n} |Y_{t_k}^m - Y_{t_k}| \to 0$ *in probability.*

More details can be found in [5].

3 Stochastic Representation of Solution to the Heat Equation

Before presenting a general method, we will illustrate our ideas by constructing a model which is equivalent to a commonly used filter, namely, the convolution of the noise image with two-dimensional Gaussian mask. The definition of our model is based on the construction of the processes X and Y. We suppose for a while that the image is given by a function defined on the whole plane. Put

$$\begin{cases} X_t = W_t^x, & t \in [0, T], \\ Y_t = u_0(X_T) - \int_t^T Z_s\, dW_s, & t \in [0, T], \end{cases} \tag{5}$$

where W^x is a Wiener process starting from $x \in \overline{D}$. From the equation (5) we can deduce that

$$\begin{cases} X_t = W_t^x, \quad t \in [0, T], \\ Y_0 = u_0(X_T) - \int_0^T Z_s\, dW_s = \mathbf{E}\left(u_0(X_T)\right) = \int_{\mathbf{R}^2} G_{\sqrt{T}}(x - y) u_0(y)\, dy, \end{cases} \tag{6}$$

where $G_\sigma(x) = \frac{1}{2\pi\sigma^2} e^{-\frac{|x|^2}{2\sigma^2}}$ is the two-dimensional Gaussian mask.

A value of the process Y at time $t = 0$ is the reconstructed pixel $u(x)$. Therefore, by (6) the image is the convolution of the noisy image with two-dimensional Gaussian mask.

While discussing the above example, we assumed that the image is the function given on the whole plane. Since we want to consider the image as a function defined on the bounded, convex set, we have to introduce a new assumption for the process X. We assume that the process X is a stochastic process with reflection with values in $\overline{D}$. In this case process X is a Wiener process with reflection, which we can write as

$$\begin{cases} X_t = W_t^x + K_t^{\overline{D}}, & t \in [0,T], \\ Y_t = u_0(X_T) - \int_t^T Z_s \, dW_s, & t \in [0,T]. \end{cases} \tag{7}$$

4 Smoothing Filters

In the case of smoothing images we will consider BSDEs associated with $\xi = u_0(X_T)$ and $f(t,y) = 0$, where X is a diffusion process with reflection.

4.1 Anisotropic Diffusion

The model constructed in the previous section is equivalent to the convolution of the noisy image with two-dimensional Gaussian mask. This filter removes noise and blurs edges. In this section we follow by [7] and provide a construction with the following properties: noise is removed and image has sharp edges.

In a neighbourhood of an edge, the image exhibits a strong gradient. In order to preserve this edge, we should diffuse along it. At locations where the variations of the brightness are weak (low gradient), we would like to encourage smoothing, the same in all directions. These conditions may be achieved by imposing

$$X_t = x + \int_0^t \sigma(X_s) \, dW_s + K_t^{\overline{D}}, \tag{8}$$

where

$$\sigma(X_t) = \begin{bmatrix} -\lambda_1(|\nabla h(X_t)|) \frac{h_{x_2}(X_t)}{|\nabla h(X_t)|}, \lambda_2(|\nabla h(X_t)|) \frac{h_{x_1}(X_t)}{|\nabla h(X_t)|} \\ \\ \lambda_1(|\nabla h(X_t)|) \frac{h_{x_1}(X_t)}{|\nabla h(X_t)|}, \lambda_2(|\nabla h(X_t)|) \frac{h_{x_2}(X_t)}{|\nabla h(X_t)|} \end{bmatrix}, \quad h = G_\gamma * u_0,$$

$$\lim_{s\to 0} \lambda_1(s) > 0, \lim_{s\to 0} \lambda_2(s) > 0, \lim_{s\to 0} \frac{\lambda_1(s)}{\lambda_2(s)} = 1, \lim_{s\to +\infty} \lambda_1(s) > 0, \lim_{s\to +\infty} \lambda_2(s) = 0.$$

To avoid false detections due to noise, u_0 is convolved with a Gaussian kernel G_γ (in practice 3×3 Gaussian mask). As an example λ_1 and λ_2 we can use functions which are shown in Fig. 1 a).

Some results from our evaluation experiments regarding the smoothing BSDE method and classic Perona-Malik method are presented in Fig. 2. It can be observed in Fig. 2 c) d) that the image created by the BSDE method is visually more pleasant than Perona-Malik algorithm. The PDE method show clear evidence of a block image. The stair-case effect is reduced in the new algorithm.

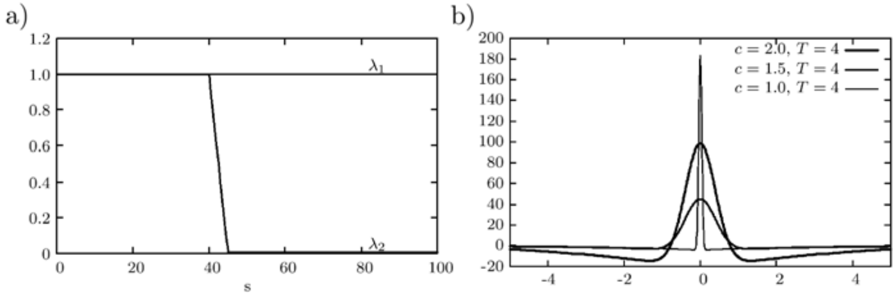


Fig. 1. a) Functions λ_1 and λ_2 b) Function MASK for different values of parameter c

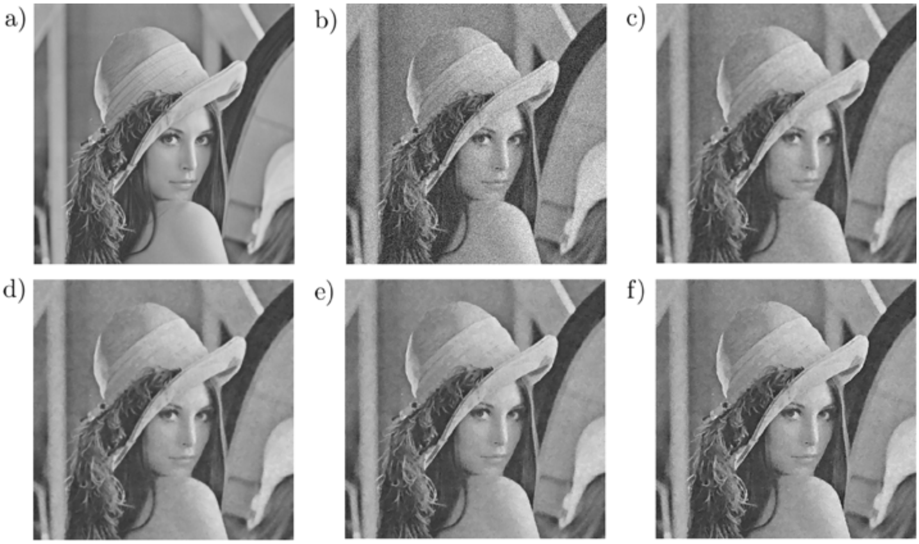


Fig. 2. a) Original image: 512 × 512 pixels b) Noisy image: $\rho = 40$ c) Perona-Malik d) Smoothing BSDE $T = 18$ e) Enhancing BSDE $T = 1.5, c = 1$ f) Enhancing BSDE $T = 1.5, c = 1.5$

5 Enhancing Filters

In the case of enhancing images we will consider BSDEs associated with $\xi = u_0(x)$ and $f(t, y) = c(y - u_0(X_t))$, where X is a Wiener process with reflection and $c \in \mathbf{R}_+$ is some constant.

Theorem 3. *Let $u_0 : \mathbf{R}^2 \to \mathbf{R}$, $x \in \mathbf{R}^2$, $f(t, y) = c(y - u_0(W_t^x))$, where W^x is two-dimensional Wiener process starting from x, $c > 0$. If (Y, Z) is a solution to BSDE associated with $\xi = u_0(x)$ and f then*

$$\lim_{m \to +\infty} Y_0^m = Y_0, \tag{9}$$

where

$$Y_0^m = \sum_{k=0}^{m-1} a_k (G_{\sqrt{t_k}} * u_0)(x), \tag{10}$$

$a_0 = \left(1+\frac{cT}{m}\right)^m - \frac{cT}{m}$, $a_k = -\frac{cT}{m}\left(1+\frac{cT}{m}\right)^k$, $t_k = \frac{kT}{m}$, $k = 1, 2, ..., m-1$.

Proof

$$\begin{aligned}
Y_{t_m}^m &= u_0(x), \\
Y_{t_{m-1}}^m &= \left(1+\frac{cT}{m}\right)\mathbf{E}\left[u_0(x)|\mathcal{F}_{t_{m-1}}^m\right] - \frac{cT}{m}u_0(W_{t_{m-1}}^x) \\
Y_{t_{m-2}}^m &= \left(1+\frac{cT}{m}\right)\left(1+\frac{cT}{m}\right)\mathbf{E}\left[u_0(x)|\mathcal{F}_{t_{m-2}}^m\right] + \\
&\quad -\frac{cT}{m}\left(1+\frac{cT}{m}\right)\mathbf{E}\left[u_0(W_{t_{m-1}}^x)|\mathcal{F}_{t_{m-2}}^m\right] - \frac{cT}{m}u_0(W_{t_{m-2}}^x) \\
&\ \vdots \\
Y_{t_{m-m}}^m &= \left(1+\frac{cT}{m}\right)^m \mathbf{E}u_0(x) - \sum_{k=1}^{m-1}\frac{cT}{m}\left(1+\frac{cT}{m}\right)^k \mathbf{E}u_0(W_{t_{m-k}}^x) - \frac{cT}{m}u_0(W_{t_{m-m}}^x)
\end{aligned}$$

It should be observed that Y_0^m is deterministic (Y_0^m is $\mathcal{F}_0^m$ measurable), $W_{t_0}^x = x$ and $\mathbf{E}u_0(W_t^x) = (G_{\sqrt{t}} * u_0)(x)$. □

Remark 1

$$\sum_{k=0}^{m-1} a_k = 1 \tag{11}$$

Proof

$$\begin{aligned}
\sum_{k=0}^{m-1} a_k &= \left(1+\frac{cT}{m}\right)^m - \frac{cT}{m} - \frac{cT}{m}\sum_{k=1}^{m-1}\left(1+\frac{cT}{m}\right)^k \\
&= \left(1+\frac{cT}{m}\right)^m - \frac{cT}{m} - \frac{cT}{m}\frac{\left(1+\frac{cT}{m}\right)\left(1-\left(1+\frac{cT}{m}\right)^{m-1}\right)}{1-\left(1+\frac{cT}{m}\right)} \\
&= \left(1+\frac{cT}{m}\right)^m - \frac{cT}{m} + \left(1+\frac{cT}{m}\right) - \left(1+\frac{cT}{m}\right)^m = 1.
\end{aligned}$$
□

Recall that the reconstructed pixel is the value Y_0^m, which can be rewritten as

$$Y_0^m = \sum_{k=0}^{m-1} a_k (G_{\sqrt{t_k}} * u_0)(x) = \left(\sum_{k=0}^{m-1} a_k G_{\sqrt{t_k}}\right) * u_0(x) = (\mathrm{MASK} * u_0)(x). \tag{12}$$

In the picture Fig. 1 b) we can see the function MASK for different values of parameter c. If c is greater than MASK is more enhancing.

Since the image is a function defined on the bounded set we have to consider Wiener process with reflection in values in domain $\overline{D}$, i.e.

$$u(x) = Y_0^m = \sum_{k=0}^{m-1} a_k \mathbf{E}(W_{t_k}^x + K_{t_k}^{\overline{D}}). \tag{13}$$

The figure Fig. 2 e) and Fig. 2 f) presents results of the BSDE enhancing method applied to the Fig. 2 d).

From the numerical point of view we have to use Monte Carlo method to calculate $u(x)$ (since it does not exist straight formula to the solution to the Skorokhod problem and reflected SDE in two dimensional case).

6 Conclusions

In this paper we present a new method to the image reconstruction which is the alternative to PDE methods. It is based on advanced tools of stochastic analysis such as BSDEs. The idea presented here provides a new methodology to solve the inverse problem.

References

1. Abraham, R., Riviere, O.: Forward-backward stochastic differential equations and PDE with gradient dependent second order coefficients. ESAIM P&S 10, 184–205 (2006)
2. Aubert, G., Kornprobst, P.: Mathematical problems in image processing. Springer, New York (2002)
3. Chan, T.F., Shen, F.: Image Processing and Analysis – Variational, PDE, wavelet, and stochastic methods. SIAM, Philadelphia (2005)
4. Juan, O., Keriven, R., Postelnicu, G.: Stochastic Motion and the Level Set Method in Computer Vision: Stochastic Active Contours. Int. J. Comput. Vision 69(1), 7–25 (2006)
5. Ma, J., Protter, P., San Martín, J., Torres, S.: Numerical method for backward stochastic differential equations. Ann. Appl. Probab. 12(1), 302–316 (2002)
6. Pardoux, É.: Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic PDEs of second order. In: Stochastic Analysis and Related Topics, VI (Geilo, 1996), vol. 42, pp. 79–127 (1998)
7. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach. Intell. 12(7), 629–639 (1990)
8. Słomiński, L.: Euler's approximations of solutions of SDEs with reflecting boundary. Stoch. Proc. Appl. 94, 317–337 (2001)
9. Tanaka, H.: Stochastic differential equations with reflecting boundary condition in convex regions. Hiroshima Math. J. 9(1), 163–177 (1979)
10. Unal, G., Krim, H., Yezzi, A.: Stochastic differential equations and geometric flows. IEEE Trans. Image Process. 11(12), 1405–1416 (2002)
11. Unal, G., Ben-Arous, G., Nain, D., Shimkin, N., Tannenbaum, A., Zeitouni, O.: Algorithms for stochastic approximations of curvature flows. In: Image Processing, ICIP 2003, Proceedings, vol. 2, vol. 3, pp. 651–654 (2003)

An Analysis of Different Clustering Algorithms for ROI Detection in High Resolutions CT Lung Images

Alfonso Castro, Carmen Bóveda, Alberto Rey, and Bernardino Arcay

Faculty of Computer Science, Campus Elviña, University of A Coruña, Spain
{alfonso.castro,carmen,alberto.rey,bernardino.arcay}@udc.es

Abstract. The detection of pulmonary nodules in radiological or CT images has been widely investigated in the field of medical image analysis due to the high degree of difficulty it presents. The traditional approach is to develop a multistage CAD system that will reveal the presence or absence of nodules to the radiologist. One of the stages within this system is the detection of ROIs (regions of interest) that may possibly be nodules, in order to reduce the scope of the problem. In this article we evaluate clustering algorithms that use different classification strategies for this purpose. In order to evaluate these algorithms we used high resolution CT images from the LIDC (Lung Internet Database Consortium) database.

1 Introduction

The analysis of lung CT images has been widely investigated due to the difficulty in locating the various possible pathologies or detecting the presence of extraneous elements such as nodules, tumours, etc. The reason for this is the high degree of complexity of the lung structure.

The modality of image most commonly used for diagnostic purposes has until now been the chest X-ray, due to its low cost. This notwithstanding, there is a steady increase in the use of high resolution CT images that enable the lung structure to be examined with a high degree of definition, since image acquisition can take place within a single breath-hold, with resolutions of less than 1 mm. There is also an increasing use of multislices CT [1] which provide a more exact image of the area to be explored, but their global implementation is still reduced and they are rather expensive.

Within this field, one of the areas that has received the greatest amount of attention has been the detection of pulmonary nodules, due to the high incidence of lung cancer in modern societies, a disease with one of the highest mortality rates and in which early detection is a vital factor for a favourable outcome.

This task is a time-consuming one for the radiologist for two reasons: on the one hand the large amount of information that has to be analysed (more than 300 thin-section images), and on the other the difficulty in distinguishing

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 241–248, 2010.

nodules in their initial stages due to their lack of definition and similarity with other elements present in the lung.

In consequence, a wide variety of CAD (Computer Aided Diagnosis) systems have been developed to undertake this task, using a similarly wide range of techniques [2] [3]. Our research group has embarked upon the development of a system of this kind.

The first task that a system of this nature has to undertake is a pre-processing stage to isolate the pulmonary lobes and eliminate any external element that might affect the subsequent classification. The system that we are currently developing includes such an initial stage.

We are now focusing on the development of the next stage, whose purpose is to detect ROIs in order to reduce the search area and obtain the lowest possible number of candidate areas for the presence of nodules, with the intention of reducing the number of false positives and increasing that of true positives. To this end we have analysed clustering algorithms that employ different strategies to classify the various elements that go to make up an image, and which have already yielded good results in the classification of other medical images, albeit of a very different nature. [4]

2 Test CT Images

For the purposes of our analysis we used a set of high-resolution CT lung images from the LIDC database [5]. The aim of this project is to develop a reference repository of CT lung images for the development and evaluation of CAD systems for the detection of pulmonary nodules, five North American institutions have collaborated in its construction.

Each image is stored according to the DICOM standard, the size being 512x512 and 16 bits per pixel.

In their studies they have incorporated an XML file of annotations indicating the presence or absence of one or more nodules, their type and their outline (determined by the coordinates of the pixels that constitute it). Those represent the ground truth for each slice.

3 Algorithms Tested

For the purposes of this study we selected only fuzzy clustering algorithms since they give a better classification than traditional algorithms. [4]

Within this sub-set we looked for algorithms that represented the state-of-the-art in this field, such as algorithms which use kernel functions (KFCM) or spatial information (SFCM) to determine classification. We also studied algorithms that they provided good results with biomedical images (FKNN).

3.1 KFCM (Kernelized Fuzzy C-Means)

This algorithm was proposed by Wu Zhong-Dong, Xie Wei-Xin and Yu Jian-Ping [6], the purpose of the kernel function is to "simulate" the distances that would be obtained by transferring the points to a space with more dimensionality, which in most cases would imply exaggerated computational costs.

The most often used kernel functions are the polynomial functions and the radial base functions.

The algorithm consists of the following steps:

1. Calculation of the membership function:

$$u_{jk} = \frac{(1/d^2(x_j, V_k))^{1/(m-1)}}{\sum_{j=1}^{C}(1/d^2(x_j, V_k))^{1/(m-1)}} \ . \tag{1}$$

   where $d^2(x_j, V_k) = K(x_j, x_j) - 2K(x_j, V_k) + K(V_k, V_k)$, m is a fuzzy index that determines the fuzziness of the clusters, C is the number of clusters, V are the centroids and K is the kernelized function.
2. Calculation of the new kernel matrix $K(x_j, \hat{V}_k)$ and $K(\hat{V}_k, \hat{V}_k)$:

$$K(x_j, \hat{V}_k) = \varphi(x_j) \cdot \varphi(\hat{V}_k) = \frac{\sum_{i=1}^{N}(u_{ik})^m K(x_i, x_j)}{\sum_{i=1}^{N}(u_{ik})^m} \ . \tag{2}$$

   where $\varphi(\hat{V}_k) = \frac{\sum_{j=1}^{N}(u_{jk})^m \varphi(x_j)}{\sum_{j=1}^{N}(u_{jk})^m}$, N is the number of pixels in this case and φ is the nonlinear mapping function approximate by the kernel.
3. Update the memberships u_{jk} to $\hat{u}_{jk}$ by means of Equoation 1.
4. If the error stays below a determined threshold, stop. In the contrary case, return to step 1.

3.2 FKNN (Fuzzy K-Nearest Neigbhour)

The Fuzzy K-Nearest Neighbour [7] is, as its name indicates, a fuzzy variant of a hard segmentation algorithm. It needs to know the number of classes into which the set that must be classified will be divided.

The element that must be classified is associated to the class of the nearest sample among the K most similar ones. These K most similar samples are known as "neighbours"; if, for instance, the neighbours are classified from more to less similar, the destination class of the studied element will be the class of the neighbour that is first on the list.

We use the expression in Equation 3 to calculate the membership factors of the pixel to the considered clusters:

$$u_i(x) = \frac{\sum_{j=1}^{K} u_{ij} \left(\frac{1}{\|x - x_j\|^{\frac{2}{m-1}}} \right)}{\sum_{j=1}^{K} \left(\frac{1}{\|x - x_j\|^{\frac{2}{m-1}}} \right)} \ . \tag{3}$$

where u_{ij} represents the membership factor of the $j - th$ sample to the $i - th$ class; x_j represents one of the K samples that are most similar to the treated pixel; x represents the pixel itself; m is a weight factor of the distance between the pixel and the samples and $u_i(x)$ represents the level of membership of the pixel x to class i.

3.3 SFCM (Spatial Fuzzy C-Means)

This algorithm [8] uses a spatial function which is the summation of the membership function in the neighborhood of each pixel under consideration. The main advantages deriving from the use of a spatial function is the possibility of obtaining more homogeneous regions and less sensitivity to noise.

In the initial stage the algorithm applies the traditional FCM (Fuzzy C-Means) [9] algorithm to obtain the initial memberships for each pixel, the iterative stage being omitted. It then calculates the spatial function value for each pixel in the image:

$$h_{ij} = \sum_{k \in NB(x_j)} u_{ik} \quad . \tag{4}$$

Where $NB(x_j)$ represents a square window centred around the pixel under consideration, its size being a configurable parameter of the algorithm. The more neighbouring pixels that belong to the same cluster, the higher the value of the function.

The next step is to calculate the spatial membership function:

$$u'_{ij} = \frac{u_{ij}^p h_{ij}^q}{\sum_{k=1}^{C} u_{kj}^p h_{kj}^q} \quad . \tag{5}$$

Where p and q are control parameters for the importance of functions u_{ij} and h_{ij}. Finally, the new centroids are calculated:

$$c_j = \frac{\sum_{i=1}^{N} u'^m_{ij} x_i}{\sum_{i=1}^{N} u'^m_{ij}} \quad . \tag{6}$$

The error is calculated, when is below a determined threshold the algorithm will stop: if not, it will recalculate the FCM in order to commence a further iteration.

4 Evaluation Metrics

For the purposes of our analysis we selected 23 CT images that represented the different types of nodules that might appear: initial stage, adherent to the pulmonary membrane, etc., and taken from three different thoracic zones: upper, middle and lower.

At that stage in the system development, we preferred to analyse a rather low number of images in order to better determine the different features that affect the result and detect possible problems.

In order to measure the quality of the results obtained we calculated the true positives (TP) and false positives (FP). This metric was chosen because the masks supplied by the LIDC only mark the zones that constitute a nodule, no information being given about the other elements that appear in the CT image, and the construction of our own masks with these zones also marked is a task that exceeds the possibilities of our group.

Although this metric does not provide as much information as others, it should be sufficient to give us a good idea of the quality of the results obtained by the different algorithms: if they are good at detecting ROIs (areas that may be nodules), then on the basis of the masks used we would get a high number of true positives and a low number of false positives.

The results are represented in the form of charts that will enable us to see under what conditions the best results for each algorithm are obtained, the aim being to get an idea of the stability of the algorithms analysed. Each chart shows the obtained TP and FP and as such summarizes the behaviour of each algorithm in the global images set.

5 Results

The main obstacle encountered when comparing the algorithms was the difference between their parameters, which made it difficult to establish any direct comparison between them, so we therefore decided to base our analysis on the shared parameters alone: the number of clusters, after which we varied the samples supplied to the algorithms in order to ascertain their sensitivity, this being the most complex parameter to determine. Finally, we carried out an analysis of each of the parameters specific to each algorithm.

The first algorithm to be analysed was the KFCM, varying the number of clusters varied from 3 to 5, with 3 sets of samples, m varied between [1.5, 3.5] and the sigma of kernel function was assigned the values: 5, 100, 150, 200 and 300.

An analysis of the results for this algorithm showed a fairly high degree of success when the number of clusters is low, but also a high number of false positives. When the number of clusters increases the degree of success is lower as a result of the assignation of points, mainly on the edge of the nodule, to another cluster. We can also see that the degree of success decreases for a high number of clusters, this in our opinion being due to the difficulty in selecting high quality samples and the notable sensitivity of this algorithm to this parameter.

In Figure 1 we can see the values obtained for one of the images used in the analysis in percentage terms. The first row of figures corresponds to the value of sigma, the second to that of m and the third is the number of clusters.

The same procedure was repeated for the FKNN algorithm, the number of clusters was fixed to 3 and 4, the set of samples used and the number of number of neighbours. The results are shown in Figure 2.

This algorithm gave good results, but at the expense of a very high number of false positives in many images. The reason is that we have to select a set of samples that clearly represent each cluster (element) that we wish to detect, which in turn results in few errors for certain images but also very large amounts of errors (approximately 50%) for others. This reduces the usefulness of the results, and makes it almost impossible to find a set of samples that provides acceptable results for larger CT sets.

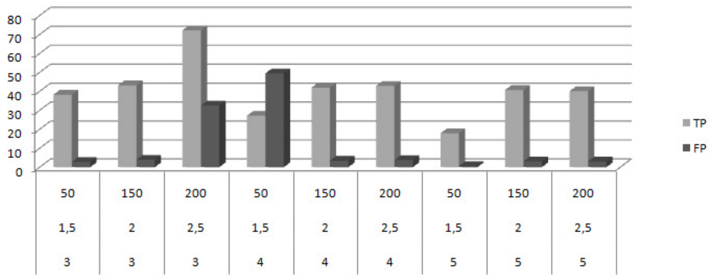


Fig. 1. Results obtained for the KFCM algorithm for a CT image used in the study

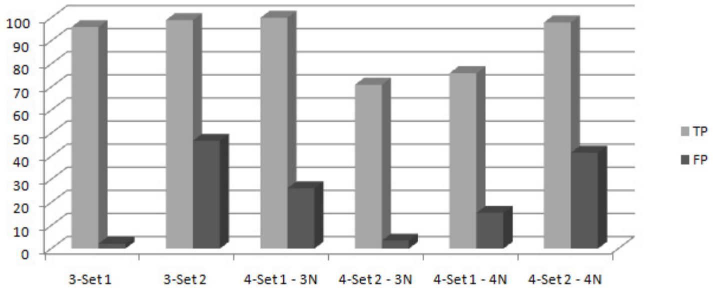


Fig. 2. Results obtained for the FKNN algorithm

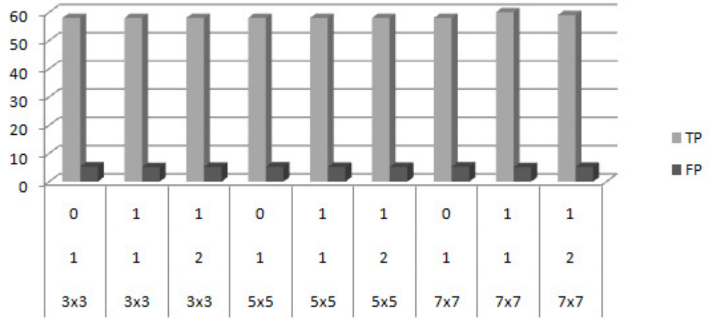


Fig. 3. Results obtained for the SFCM algorithm for a CT image

The next algorithm to be analysed was the SFCM, for which we varied the following parameters: number of clusters, size of the neighboured, samples, m, p and q. Figure 3 shows the results obtained: the first row of figures corresponds to p, the second to q and the third to the size of the window.

It is clear from the chart that the degree of success is not particularly high, but that the number of false positives is low, in this case with four clusters.

The reason why a greater degree of success was not obtained for a high number of clusters is, in our opinion, the same as for the KFCM algorithm. It should also be noted that the most critical parameter in this algorithm is window size, rather than any of the weighting factors.

5.1 Conclusions

The testing results show good success rates for certain algorithms (between 70-80%), but with a considerable amount of false positives: in many cases, the lung tissue is difficult to distinguish because of the value of the nodule pixel, which is used as classification basis. The Figure 4 shows one of the best results obtained and worst for the 23 slices set.

On the other hand, in most cases the false negatives can be recuperated by means of region growth techniques, using the detected node area as seed.

We believe that the stability provided by the algorithms, i.e. the success rate, remains stable in spite of the variety in samples; their fundamental parameters for KFCM and SFCM can serve as a basis for the detection of ROIs if the classification can become more based on homogeneity and the combination of different strategies.

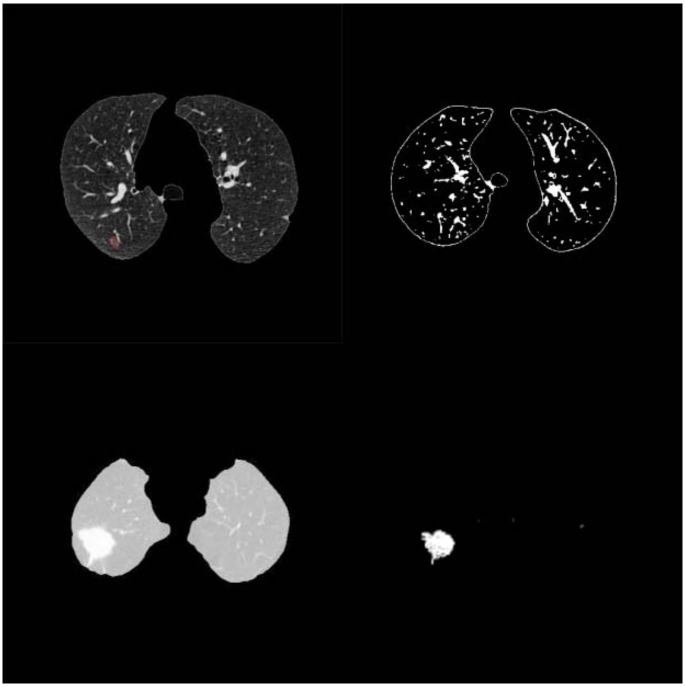


Fig. 4. One of the worst and the best results obtained, in this case using KFCM algortihm

We observed that most node pixels present homogeneous values and are currently searching for mechanisms that allow us to incorporate this factor into the classification. We are initiating the study of assignation to clusters according to the membership of various clusters rather than the degree of membership, as was the case until now. We are also combining the above strategies in order to improve the results.

Acknowledgments. This work has been supported by the project 09SIN006PR from I+D Suma of Plan Galego de Investigación, Desenvolvemento e Innovación Tecnolóxica (INCITE), Xunta de Galicia.

References

1. Hwang, J., Chung, M., Bae, Y., Shin, K., Jeong, S.Y., Lee, K.S.: Computer-Aided Detection of Lung Nodules: Influence of the Image Reconstruction Kernel for Computer-Aided Detection Performance. Journal of Computer Assisted Tomography 34(1), 31–34 (2010)
2. Lee, Y., Hara, T., Fujita, H., Itoh, S., Ishigaki, T.: Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique. IEEE Transactions on Medical Imaging 27(7), 595–604 (2001)
3. Matsumuto, S., Kundel, H.L., Gee, J.C., Gefter, W.B., Hatabu, H.: Pulmonary nodule detection in CT images with quantized convergence index filter. Medical Image Analysis 10, 343–352 (2006)
4. Castro, A., Bóveda, C., Arcay, B.: Analysis of Fuzzy Clustering Algorithms for the Segmentation of Burn Wounds Photographs. In: Campilho, A., Kamel, M.S. (eds.) ICIAR 2006. LNCS, vol. 4142, pp. 491–501. Springer, Heidelberg (2006)
5. Armato III, S.G., McLennan, G., McNitt-Gray, M.F.: For the Lung Image Database Consortium Research Group.: Lung Image Database Consortium: Developing a Resource for the Medical Imaging Research Community. Radiology 232, 739–748 (2004)
6. Zhong, W.D., Wei, X.X., Jian, Y.P.: Fuzzy C-Means clustering algorithm based on kernel method. In: Proceedings of the Fifth International Conference on Computational Intelligence and Multimedia Applications, ICCIMA 2003 (2003)
7. Bandyopadhyay, S., Maulik, U.: Genetic Clustering for Automatic Evolution of Clusters and Application to Image Classification. Pattern Recognition 35, 1197–1208 (2002)
8. Chuang, K., Tzeng, H., Chen, S., Wu, J., Chen, T.: Fuzzy C-means clustering with Spatial information for image segmentation. Computerized Medical Imaging and Graphics 30, 9–15 (2006)
9. Bezdek, J.C.: Fuzzy Mathematics in Pattern Classification. Ph. D. dissertation, Appl. Math., Cornell University, Ithaca, NY (1973)

A System to Measure Gap Distance between Two Vehicles Using License Plate Character Height

KinYip Chan, Andrzej Ordys, and Olga Duran

Faculty of Engineering, Kingston University, London, UK, SW15 3DW
k0963725@kingston.ac.uk

Abstract. This paper describes a vision-based car distance measuring system capable of telling the driver the gap distance between the host vehicle and the vehicle in front. The aim is to increase the road safety by warning the driver if the driving distance is too close and therefore can cause dangerous situation, and hence provide comfort driving condition for car users. The system uses the size of number plate characters to determine the distance. With the help of an image preprocessing stage, the region of interest (ROI) in the acquired images is identified. The ROI is then examined by a rule-based algorithm that identifies the characters in the plate and computes the corresponding height of the plate characters and thus the distance between the cars. Finally, in order to reduce the complexity of the algorithm, we propose a number plate tracking technique that continuously tracks and computes the height of the characters. We show the system working in real situations and results are discussed.

1 Introduction

Technology has increasingly been used to improve road safety. The so-called Collision Avoidance and Collision Warning systems are being developed by major automotive manufactures. Such systems can adaptively act as a co-driver for a vehicle user, therefore assuming control of the vehicle brakes and steering, or act as an advisor, giving warnings in a hazardous situation [1][2]. In current automotive research, companies mainly focus on two methodologies in collision detection using sensors - radar-based [3] and vision-based [4]. Radar-based collision detection has a transceiver installed in front of the host vehicle. It transmits waves, such as infra-red or ultrasonic, receives the reflections. The system calculates the reflected wavelength and hence finds the distance between the host and the target vehicle. The radar-based systems have limitations such as: being unable to detect whether the target is moving or stationary, also the angle of the infra-red beam is narrower over longer distances. The other issue is the bandwidth and regulatory compliance of the transceiver. Currently vision based systems use the base of rear chassis width of a vehicle to determine the driving gap distance for Adaptive Cruise Control (ACC) [4]. This paper explains the

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 249–256, 2010.

research on using number plate character size to determine the driving gap distance. The camera finds the location of the number plate [5] on the preceding vehicle and, through image processing, the size of symbols on that number plate will be determined. As number plates use standardized fonts, this size is related to the distance between the camera and the preceding vehicle. In contrast to existing radar systems, the proposed system will discriminate the objects which are stationary and non vehicles, and will allow unique identification of the vehicle in front if necessary. It has higher accuracy for smaller distance against radar systems. Furthermore, by tracking vehicles in adjacent lines, it enables prediction of traffic behavior, and can act as a co-driver.There is some work done in the field of number plate localisation and identification [6][7][8][9][10], but to the best knowledge of the authors, the size of the characters has not been used to determine the driving gap distance so far. This paper is organized as follow. The system structure and requirements are described in Section 2. Section 3 describes the algorithm. The implementation is shown in Section 4. Results are presented in Section 6. Finally, conclusions are drawn in Section 7.

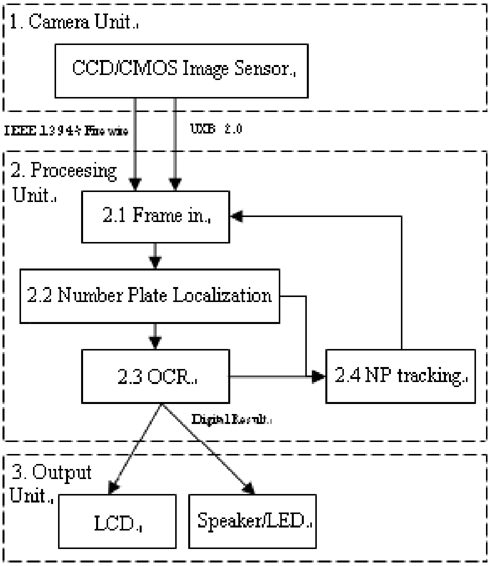


Fig. 1. General overview of the algorithm

2 Methodology

The systems consist of three units, namely the camera, the processing and output units (Fig. 1). The scanned image is then sent to the processing unit where processing is divided into two stages, namely number plate localization and number plate tracking (Fig. 2). Those will be describe in detail in the following section. Finally the processed images together with the computed distance are displayed in the output unit.

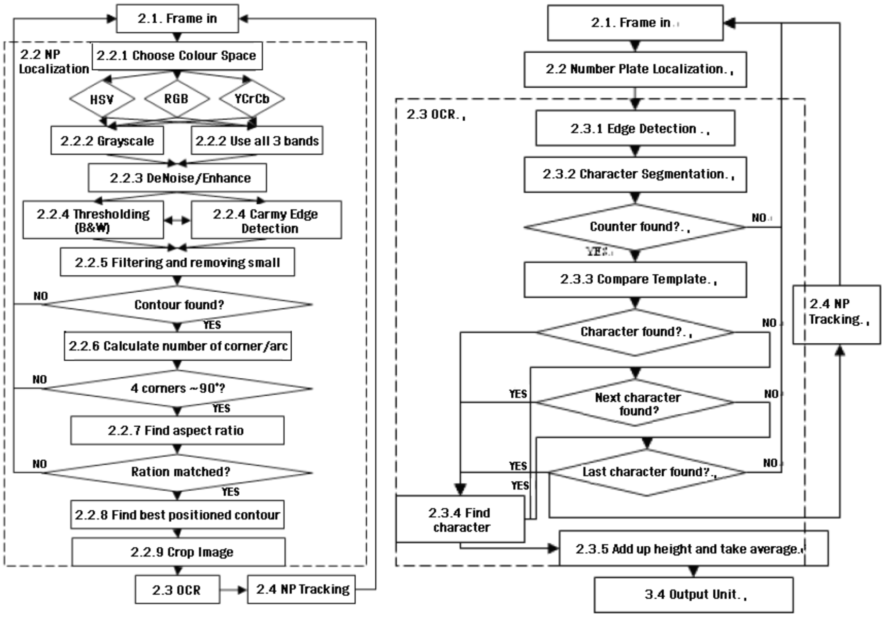


Fig. 2. Number plate localization, optical character recognition (OCR) unit blocks diagrams

2.1 Number Plate Localization

Number plate Localization blocks acquire the images from the camera. By using an image processing procedure, the number plate is segmented, and the cropped image is sent to OCR block for character recognition [11][12][13][5]. The sequence of operations for number plate location is as follows.

Select colour bands (see 2.2.1 in Fig 2.a). Different colour bands can be chosen. RGB, HSV, YCrCb. Transform colour band into monochrome band, for extracting number plate using 2D algorithms (2.2.2). There are two ways; either using 3 bands separately, in this case no processing is performed, or converting into grayscale by weighting. The example for grayscale representation of its luminance of RGB is typically:

$$Y = 0.3 * R + 0.59 * G + 0.11 * B$$

De-noise 2D images using Gaussian or median filter (2.2.3). The image can be enhanced using dilate and erode method to sharpen the edges [5][7]. Extract the contours using edge detection (for three components) or thresholding or binarisation [14] for grayscale case (2.2.4). For edge detection, canny edge detector is used as it has been tested to outperform other algorithms, such as Sobel or Hewitt. Filter out any contour which is too small. Find contours with 4 corners such that all angles are close to 90 degrees (2.2.6). Calculate the height/width ratios of contours (2.2.7) and those rectangles which aspect ratio does not match with standard number plate sizes are filtered out. From the rectangular contours left on the image (2.2.8) - number plate candidates -find the one closest to the middle of x-axis. The data is saved for number plate tracking and the selected rectangle is cropped for OCR use.

2.2 Optical Character Recognition (OCR)

The OCR block is expanded in Fig. 2.b. It takes as input the cropped image from number plate localization block. Next, the following operations are performed: The pre-processing changes the number plate image into black and white (See 2.3.1 in Fig 3) by using thresholding or edge detection as in (2.2.4 in 2). Label and isolate the character contours (2.3.2). Perform OCR by comparing the contours with stored templates (2.3.3). If there is no character found in the contour, the next contour is processed. If a character is found the height of that contour is recorded (Fig7c) (2.3.4). To increase the accuracy, all the contours found on the number plate are processed with OCR and the average is taken (2.3.5). However, it is worth noting that it is not important to recognize all the characters on the number plate or to recognize any of the characters correctly. The averaged height is directly related to the gap distance between the camera and the front vehicle. Output the result by inform driver using LCD, speaker or L.E.D (3.4) [15].

2.3 Number Plate Tracking

After the number plate is found, the area of interest on subsequent pictures is reduced to a neighborhood of that number plate (Fig. 3). The advantage is that unwanted data is reduced and the processing time is reduced. The scanned image from the camera is still sent as a full frame but before processing the program resets pointer to the area of interest. The program is retaining the information about number plate position on subsequent images. A rule-based algorithm computes the position of the number plate in subsequent frames. In summary, a subimage (ROI) is searched for characters taking into account a margin of safety. If not characters are found, then the whole image is search as shown in Section 2.1 (see Fig. 2.c).

3 Results

For the experiments presented in this section, we used a Sensata firewire camera with 640X480 resolutions with RGB colour output. A 75mm focal length lens is mounted on the camera with aperture from f/3.9 to f/32. The camera connected to a evaluation box, which output the signal to laptop via firewire. The camera is mounted on the dashboard. The measured distance is in a digital format. It is sent to an output device, which can be an LCD screen, a numeric display (with a warning signal if the distance is shorter than the minimum braking distance) or the vehicle computer system where it can be processed for further use, for instance in adaptive cruise control, "stop and go" or the collision warning system.

Preliminary tests in still situation, by using progressive scanning camera showed that the accuracy and detecting rate over 80% [7] could be achieved for distances up to 45 metres. Here, we perform experiments in real driving tests under different weather conditions. Two aspects for error are analysed. On the one hand, detective error occurs when the system fails to detect the correct location of the number plate.

The detecting rate depends on the image quality and evidence of the target number plate. The program is set to reduce certain noise, but if the number plate is obstructed by dirt or shadow or if a part of the number plate is being covered, the detection will fail. If failed sample is done by poor focus or poor quality, the system continues to search for the next frame. If the number plate is obstructed badly then there is no method to compromise [10]. The detection error can be calculated by: $DetectionError = 1 - \frac{PlateChracterDetected}{TotalPlatesExist}$

The other detective error occurs when the system detects false positives (e.g. number plate is wrongly detected). It can be detected as follow:

- when characters size do not match the ratio.
- when more than the maximum number of characters allowed in plates are found (e.g. 8 in the UK).

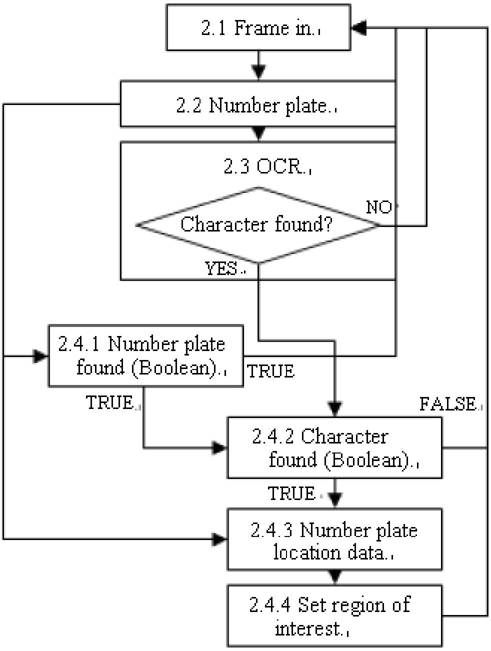

2.1 Frame in.
2.2 Number plate.
2.3 OCR.
Character found?
NO
YES.
2.4.1 Number plate found (Boolean).
TRUE
TRUE.
FALSE.
2.4.2 Character found (Boolean).
TRUE.
2.4.3 Number plate location data.
2.4.4 Set region of interest.

Fig. 3. Number plate tracking unit block diagram

On the other hand, measurement error gives the uncertainty of the result. The characteristic between character heights against distance is an exponential decline. Error increases exponentially when the gap distance is further. Measurement error also depends on weather condition. Measurement error is measured as $MeasurementError = 1 - \frac{AveragedCharacterHeight}{Meanvalueof1000samples}$.

3.1 Results Analysis

Figures 4 and 5 show the results corresponding to different weather conditions. In Tables 1and 1, we show the detection rate for different image enhancement and colour segmentation approaches. By using HSV colour space and dilate/erode and Gaussian blur before the edge detection, the best results were obtained for all testes weather conditions. In sunny, cloudy and at night we reached detection rates of over 80% with the target vehicle being at distances in the range 20-30 metres. In daytime, the detection accuracy can depend on the light source, which can affect the saturation of the image. In this test, the best results were obtained at night with the number plate outlined by the illumination from the number plate light, and hence image saturation was improved. For the rainy days, the detection rate dropped to 23%. The wipers affected some images but the main course of low detection rate was the fact that the image was not clear enough for

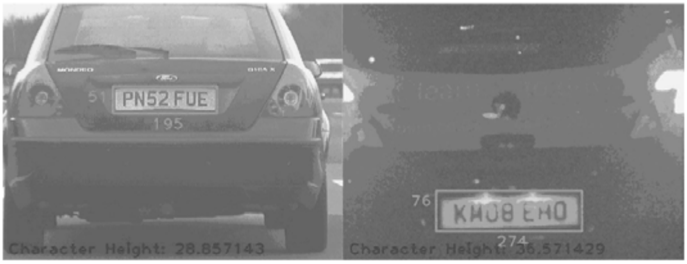


Fig. 4. Real driving condition in day and night time, respectively

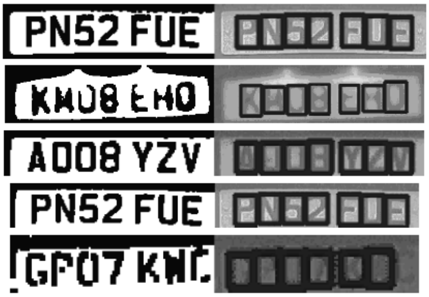


Fig. 5. Number plate cropped and characters detected during day time, night time, sunny and cloudy condition, respectively

Table 1. Detection rate comparison for Gaussian blur and dilate/erode image enhancement and colour segmentation when using the HSV colour space

Weather	Gaussian blur and dilate/erode	without Gaussian blur	without dilate/erode
Sunny	90%	87%	86%
Cloudy	85%	75%	82%
Night	93%	77%	39%
Rainy	23%	27%	3%

Table 2. As in Table 1 but using the RGB colour space

Weather	Gaussian blur and dilate/erode	without Gaussian blur	without dilate/erode
Sunny	0%	0%	32%
Cloudy	59%	60%	59%
Night	38%	34%	14%
Rainy	2%	2%	1%

the OCR. Nevertheless, HSV colour space with dilate/erode and Gaussian blur setting has been found to provide the best results. This configuration is further analysed in Table 3. It is noted that the false plate and character rate increases significantly, at 29% at night time. In Fig. 5.b, we observe that although at night time all characters can be detected, the OCR might not get the whole character height. Therefore it caused the error and the high percentage difference.

Table 3. Error with HSV colour space using Gaussian blur and dilate/erode

		Maximum height difference between characters	
Weather	False plate/ character	[pixels]	Percentage
Sunny	2.22%	3	12.5%
Cloudy	3.52%	3	10.7%
Night	7.53%	9	29.0%
Rainy	0%	4	26.67%

4 Conclusions

This article presents an innovative idea to determine the gap distance between two vehicles based on the characters height of the number plate in the car in front. An algorithm using image processing tools and a rule-based system to compute the height of the characters has been presented. The algorithm has three main stages namely, the car plate location, tracking and optical character recognition. The equipment has been installed to a vehicle to perform tests with this algorithm in real-time drive situation and very promising results have been found with detection rate in the 90% for good weather condtion. An existing vision based vehicle distance measuring system, described in [4], targets the width of the base of the vehicle chassis. For this system, the accuracy quoted is 30%. Our system compares favourably with this in all conditions, except for rainy weather or dirty plates[16]. In general, the accuracy of vision based systems is still worse than that for radar based or ultrasonic based systems but vision based systems overcome some limitations of other systems, as explained in the introduction.

References

1. ReportLinker: Driver assistance systems market overview, technical review and forecasts to 2013. technical reports (March 2007)
2. Malec, J., Morin, M., Palmavist, U.: Driver support in intelligent autonomous cruise control. In: Proceedings of the IEEE Intelligent Vehicles Symposium, pp. 160–164 (1994)
3. Gorjestani, A., Donath, M., Alexander, L.: Radar based longitudinal virtual bumper collision avoidance system implemented on a truck. Technical reports (1999)
4. Stein, G., Mano, O., Shashua, A.: Vision-based acc with a single camera: bounds on range and range rate accuracy. In: IEEE Intelligent Vehicles Symposium (IV 2003), pp. 120–125 (June 2003)
5. Hou, P., Zhao, J., Lui, M.: A license plate locating method based on tophat-bothat changing and line scanning. Journal of Physics: Conference Series (2006)
6. Draghici, S.: A neural network based artificial vision based system for license plate recognition. In: IJNS, February 1997, pp. 113–126 (1997)
7. Rezaie, F.F.A., Ziaratban, M.: A morphological-based license plate location. In: Conference, I.I., ed.: Journal. Vol. 1, 57–60 (October 2007)
8. Reginaldo, L., Curado, A.F., Bauchspiess, A., Curado, V.B.: Using pca to recognize characters in vehicle license plates. In: IPCV 2006, pp. 264–268 (2006)
9. Enyedi, B., Konyha, L., Fazekas, K.: Real time number plate localization algorithms. Journal of Electrical Engineering 57(2), 69–77 (2006)
10. Suresh, K., Kumar, G.M., Rajagopalan, A.N.: Superresolution of license plates in real traffic videos. IEEE Transactions on Intelligent Transportation 8(2) (June 2007)
11. Parker, J., Federl, P.: An approach to licence plate recognition. Technical reports (1996)
12. Zhu, W.: Number plate recognition system. Tatung University Technical report (2007)
13. Lienhart, R., Effelsberg, W.: Automatic text segmentation and text recognition for video indexing. Technocal reports 8(1), 69–81 (1998)
14. Nishi, T., Yamazaki, N., Koite, S., Kuno, T., Umezaki, T.: Collision avoidance system using laser beam. Image and Vision Computing (2005)
15. Mirmehdi, M., Clark, P., Lam, J.: Extracting low resolution text with an active camera for ocr. In: Proc. IX Spanish Sym. Pat. Rec. Image Proc., pp. 43–48 (2001)
16. BBCnews: Warning over dirty number plates (November 12, 2005)

Efficient Neural Models for Visual Attention

Sylvain Chevallier, Nicolas Cuperlier, and Philippe Gaussier

ETIS - Neurocybernetic team
ENSEA - University Cergy-Pontoise - CNRS UMR 8051
F-95000 Cergy, France
firstname.name@ensea.fr

Abstract. Human vision rely on attention to select only a few regions to process and thus reduce the complexity and the processing time of visual task. Artificial vision systems can benefit from a bio-inspired attentional process relying on neural models. In such applications, what is the most efficient neural model: spiked-based or frequency-based? We propose an evaluation of both neural model, in term of complexity and quality of results (on artificial and natural images).

1 Introduction

Biological inspiration aims at adapting biological mechanisms to design efficient artificial systems benefiting from the natural solutions. An essential mechanism of the human visual system is visual attention, which allows higher cognitive processes (such as learning or recognition) to concentrate on few regions of the visual scene, selected by the attention. Adapting attention in artificial systems may be a way to reduce the computational cost of visual tasks [10]. There is a large number of applications for such artificial attentional systems, e.g. driver assistance [15], retinal prostheses [17] or robotics [9]. Several artificial systems propose an adaptation of attentional process on a neural level, i.e. biologically plausible efficient artificial systems implemented with neural networks [1,4,6,14,19]. Based on the time scale of the description, one can distinguish two ways of representing encoded information in neural models. In spiking neuron networks (SNN), information is encoded and exchanged between neurons with spikes, i.e. transient voltage pulses. With frequency-based neural network (FNN), information is encoded in the activation of each neuron, a quantity expressing the mean frequency of spiking rate [13]. The choice of the model type has a major influence both on the computational cost of the system and on its possibilities, i.e. mechanisms which can be adapted from biological observations. Which neural model, between SNN and FNN, is the best suited to implement an efficient bio-inspired attentional system? We propose a comparison of the computational complexity of these two type of networks and an evaluation of their performances on artificial and natural images.

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 257–264, 2010.

The different psychological theories of the human attention, such as [7,18,20], agree that several local visual features (e.g. color, orientation, contrast or movement) are detected in a parallel process and then combined on a saliency map. This saliency map indicates potentially interesting regions, called saliencies. The attentional process then selects the most salient regions.

In bio-inspired attentional systems, the detection of visual is usually achieved with difference of Gaussians (DOG) filters to reproduce the spatial transformation of retinal ganglion cells [8] and Gabor wavelets to detect orientations, as observed in the simple cells of the primary visual cortex [11]. Systems implemented with FNN [1,14,19] use neural networks to combine features on the saliency map and to select the most important saliencies. This selection stage is realized with a Winner-Take-All (WTA) mechanism [21], which allow to select only the most salient region. In order to determine the next salient regions, an inhibition of return (IOR) mechanism is used to inhibit the previously selected regions. When relying on spiking neurons to simulate visual processing, it is possible to take advantage of the precise timing of spike firing to implement an anytime system. Anytime systems [2] may be stopped at any instant and return a response. The quality of response depends on the computation time allowed to the system: quality of response increases with the computation time. An anytime SNN for visual attention, as described in [5,4], is able to extract a number of saliencies which depends on the CPU time allowed to the SNN. The saliencies are detected in the order of their importance, e.g. the first saliency found is the most salient, hence there is no need of a WTA.

In Sect. 2, an architecture of attentional system and its implementation with FNN and SNN are described. A complexity analysis of the FNN and SNN implementation is detailed in Sect. 3.1 and a performance comparison on artificial and natural are proposed respectively in Sect. 3.2 and 3.3. Conclusions and perspectives are detailed in Sect. 4.

2 Models and Implementations

We use a a multiscale and multi-feature attentional architecture similar to architecture proposed by [10]. This attentional architecture uses local contrast of luminance, orientations and colors to extract saliencies. Figure 1 displays the organisation of this architecture, which is composed of 2D neural map. The luminance and colors of an input image are fed in Input maps. Detections of contrasts, orientations and color opponency are realized for a high and a low spatial frequencies. Local luminance contrasts are obtained with a DOG filtering. Orientation information are detected with Gabor wavelets filtering for four distinct orientations (0, $\frac{\pi}{4}$, $\frac{\pi}{2}$ and $\frac{3\pi}{4}$). Color opponency uses a combination of DOG filtering to detect red-green and blue-yellow opponency. The high and low spatial frequency information are combined on the saliency map.

The SNN implementation of the attentional architecture is described in [4] and uses Leaky Integrate-and-Fire neural model. The LIF model describes the evolution of an internal parameter V and when V exceeds a threshold ϑ, the neuron fires a spike. The LIF model is characterized by the following differential equation:

$$\begin{cases} \frac{dV}{dt} = -\lambda(V(t) - V_{\text{rest}}) + I_{\text{input}}(t), \text{ if } V < \vartheta \\ \text{else fires a spike and } V \text{ is set to } V_{\text{reset}} \end{cases} \tag{1}$$

where λ is the membrane relaxation constant and $I_{\text{input}}(t)$ is an input term. This SNN realizes an anytime neural filtering, leading to a gradual response which get closer to the convolution response as the computation time increases [3]. The early responses of neural filtering exhibit a bias which overvalue filtering responses for high luminance values. The saliency map gathers the neural filtering responses in order to determine the saliencies. Saliencies are thus obtained gradually and the most salient region are detected first. As the SNN extract salient regions already sorted in the order of their importance, there is no WTA. It is important to note that the result of the SNN depends on the simulated network time t.

In FNN implementation, frequency-based neural model are used to implement a classical convolution algorithm. It is computationally equivalent to filter the image with a classical algorithm and then convert the resulting image values in neuronal activity. When the input image is filtered for both spatial frequencies, the resulting activation are summed on the saliency map. To sort the saliencies in the order of their importance, the FNN relies on a WTA map coupled with an inhibition of return map (see Fig. 1). As realistic biological WTA have a high computation cost [21], we use a WTA implementation relying on a ARGMAX function which gives the same results as biologically plausible WTA but with a lower computational cost.

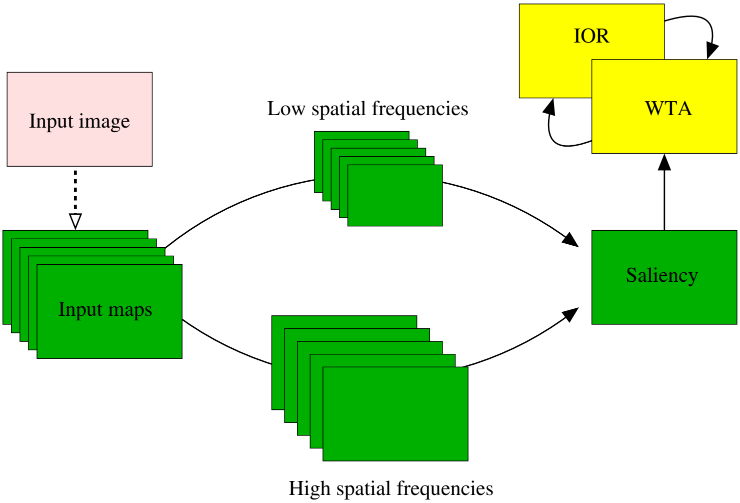


Fig. 1. Preattentive visual architecture, multiscale and multi-features (local luminance contrasts, detection of orientations and color opponency). Neural maps used in the SNN and FNN implementation are displayed in green and neural maps used only with FNN are in yellow.

3 FNN and SNN Comparisons

3.1 Complexity Analysis

The most important computational cost for FNN implementation is the image filtering cost. The image filtering is realized with a classical convolution algorithm. As the input image and the filter are relatively small, this is a acceptable choice. Hence, for an architecture processing f features at s spatial scales, with filters of size M and an input image of N pixels, the time complexity is in $\mathcal{O}(f \times s \times M \times N)$. The WTA algorithm used in this FNN as a time complexity of $\mathcal{O}(N)$. The overall time complexity is thus $\mathcal{O}(f \times s \times M \times N)$. The FNN implementation uses Promethe [12], a distributed real-time neural network simulator.

The time and space complexity of a SNN heavily depends on implementation choices. The SNN is implemented on a simulator developed by the authors, which uses a simulation strategy called hybrid-synchronous [16]. Synchronous simulators rely on an internal clock with a time step Δt to update the state variables of every neurons in the network. The choice of the time step value is crucial as it influences the computational and the precision of the obtained results. Smaller Δt value offers more precise results but higher computational cost. Here, $\Delta t = 0.1$ms which is sufficiently small to obtain precise and reproducible results. With a hybrid-synchronous strategy, only the "active" neurons are updated, i.e. neurons with non null input term $I_{\text{input}}(t)$ at instant t.

The computational cost of a simulation can be expressed as the sum of the spike propagation cost and the neuron update cost. Here, it is:

$$c_p \times F \times N \times M + c_u \times \frac{A}{\Delta t} \tag{2}$$

The total propagation cost depends on the unitary propagation cost c_p, the mean number of fired spike which depends of the mean frequency rate F and the number of neurons N and the mean number of connection per neuron (which depends on the filter size M, see [4]). The total update cost rely on the unitary update cost c_u, the mean number of active neurons A and Δt. Here the unitary update cost requires 10 FLOP.

The computational cost is dependent of the input image: a complex image (in the meaning of filter used in the SNN) induced a large number of spikes and the simulation cost is high. To illustrate this fact, we construct test images with various complexity, i.e. with different number of features. These test images are set up by converting impulse response of DOG and Gabor filter in small images (called patches). The amplitude and spatial position of these patches are drawn randomly. Left part of Fig. 2 shows a test image used to evaluate the influence of an image complexity on the required processing CPU time. CPU time (measured in CPU cycles) required to process image of growing complexity (with 1, 10, 50 or 100 patch's) is recorded and shown on the right part of Fig. 2. One can see CPU cycles needed to extract saliencies increases with the image complexity.

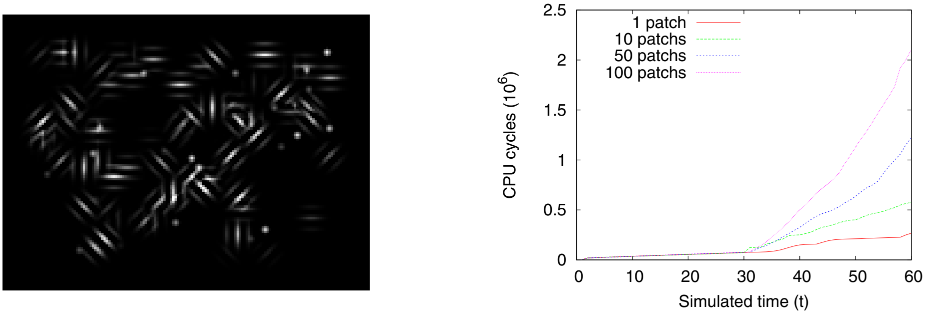


Fig. 2. Left:A test image with 100 patches. Each patch is the impulse response of a DOG or a Gabor filter. Right: CPU cycles needed to extract saliencies on images with growing complexity, i.e. different number of patches. CPU cycles are measured for each simulated time step in the SNN.

3.2 Artificial Images

We propose a comparison of saliency detected on pop-out images [18], where a target among distractors is easily identifiable. This is the case when the target differs from distractors for a given feature. Figure 3 shows the most salient region obtained on two pop-out test images.

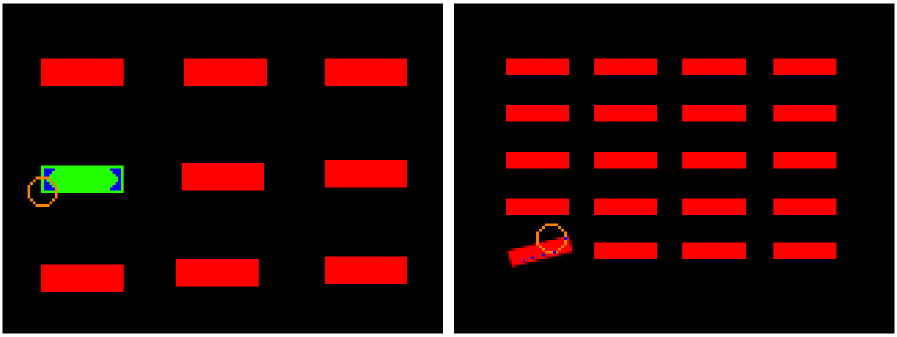


Fig. 3. Pop-out images (160x120 pixels) used for saliency detection. On the left, target differs from distractors by its color and, on the right, by its orientation. The most salient region detected by FNN is represented as an orange circle and for SNN by blue pixels.

The Figure 3 shows that both SNN and FNN are able to detect saliencies, but results take different forms. With FNN, a saliency correspond to the position of the winner neuron in the WTA, i.e. neuron with the highest activation on saliency map. As the winner location and its direct neighborhood is then inhibited by the IOR, we indicate saliency as a circle centered on the winner location. The

SNN extract salient regions already sorted in the order of their importance. The first neurons to fire on the saliency map indicate the most salient regions, so there can be several points with an identical saliency value. On the left image of Figure 3, the edges of the green target are salient and on the right image four salient dots are detected on the bottom part of the target.

3.3 Natural Images

We propose a comparison based on 19 natural images of 160x120 pixels acquired with a standard webcam. Figure 4 shows two of the 19 images and the three most salient region detected by the FNN and the SNN. On few images, salient regions are not extracted in the same order in SNN and FNN (as on the right image). These differences are due to the fact that SNN present a bias toward high luminance value. On the right image, luminance and color contrasts of the blue can (white on light background) are surevaluated compared to contrasts of cans on the bottom (light grey on dark background).

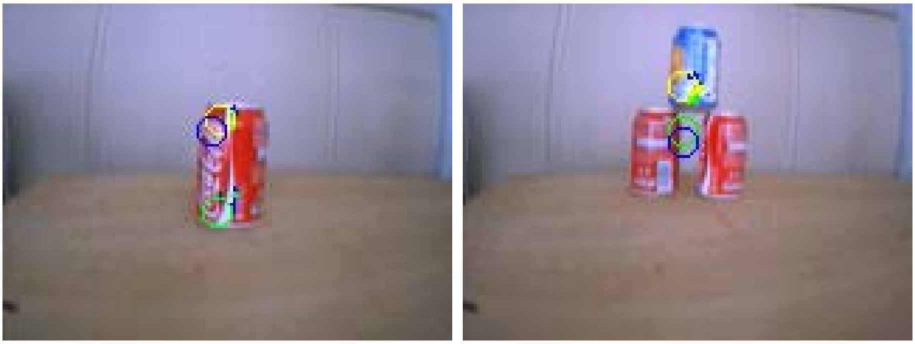


Fig. 4. Examples of saliencies obtained on natural images. For each image, the three most salient regions are respectively indicated in yellow, green and blue. FNN saliencies are indicated with circles and SNN saliencies are indicated by colored pixels.

To evaluate the computational performance of FNN and SNN, we measured the number of CPU cycles needed for each neural networks to find the most important salient region. Each measure is repeated 10 times to compute a mean number of CPU cycle and its standard deviation. CPU cycle measurements are almost constant on the 19 natural images for FNN: it required 2.68414 $\times 10^6$ CPU cycles (with a standard deviation of 0.008 $\times 10^6$) to find the most salient region. As an example, on a computer equipped with 4 AMD Opteron 2.4 GHz, the mean execution time is 0.62 sec, that is 1.6 frame per second.

The Table 1 shows the mean number of CPU cycles measured with the SNN and the difference in percent with the FNN. As previously explained, SNN uses an iterative simulation, so we chose to stop the simulation as soon as the SNN find the first salient region (usually after 45 time steps). It appears that the

Table 1. Number of CPU cycles required to find the most salient region on each of the 19 images with the SNN. For the FNN, the mean number of CPU cycles required is 2.68414 10^6. The difference between SNN and FNN is shown in the last column.

Image	SNN (10^6 CPU cycles)	Difference with FNN
1	3.389 ± 0.041	23.62 %
2	2.359 ± 0.049	**-12.74 %**
3	2.409 ± 0.006	**-11.66 %**
4	3.487 ± 0.010	28.27 %
5	3.682 ± 0.076	38.42 %
6	2.530 ± 0.006	**-3.68 %**
7	2.944 ± 0.005	12.00 %
8	2.830 ± 0.004	7.56 %
9	2.816 ± 0.004	6.01 %
10	3.336 ± 0.107	25.39 %
11	3.520 ± 0.004	32.74 %
12	2.868 ± 0.002	7.80 %
13	4.157 ± 0.006	53.07 %
14	3.994 ± 0.003	46.86 %
15	3.737 ± 0.004	35.43 %
16	4.144 ± 0.036	53.48 %
17	2.992 ± 0.097	12.46 %
18	2.348 ± 0.010	**-12.74 %**
19	2.264 ± 0.011	**-15.77 %**

mean CPU cycles required to find the most salient region varies from one image to another. This effect is due to the fact that the 19 images have different complexity, in term of the filters used in the architecture. One can see that for one fourth of the images, the SNN find the most salient region before the FNN does. For the image on the right part of Fig. 4 (denoted as image #6 in Table 1), the SNN find the most important saliency, indicated in yellow, before the FNN.

4 Conclusions

This contribution proposes a comparison of two neural model, spike-based and frequency based, to implement an artificial attentional system. FNN have a lesser computational cost than SNN but require a WTA to extract the saliencies. The SNN is an anytime system and saliencies are extracted gradually, in the order of their importance. Both neural networks indicate the same saliencies and the SNN find the first saliency before FNN in one fourth of the natural images used in the evaluation. One can note that if a FNN relies on a biologically realistic implementation of WTA, the computational cost of the FNN will be greatly increased. To implement a bio-inspired attentional system, FNN is an efficient solution. An possible solution to benefit from the advantage of both neural models is to use both FNN and SNN, running in parallel on different computers, to process visual input.

References

1. Ahrns, I., Neumann, H.: Space-variant dynamic neural fields for visual attention. In: CVPR, vol. 2, p. 318. IEEE, Los Alamitos (1999)
2. Boddy, M., Dean, T.: Deliberation scheduling for problem solving in time-constrained environments. Artificial Intelligence 67(2), 245–285 (1994)
3. Chevallier, S., Dahdouh, S.: Difference of gaussians type neural image filtering with spiking neurons. In: IJCCI, pp. 467–472 (2009)
4. Chevallier, S., Tarroux, P.: Covert attention with a spiking neural network. In: Gasteratos, A., Vincze, M., Tsotsos, J.K. (eds.) ICVS 2008. LNCS, vol. 5008, pp. 56–65. Springer, Heidelberg (2008)
5. Chevallier, S., Tarroux, P., Paugam-Moisy, H.: Saliency extraction with a distributed spiking neural network. In: ESANN pp. 209–214 (2006)
6. de Brecht, M., Saiki, J.: A neural network implementation of a saliency map model. Neural Networks 19(10), 1467–1474 (2006)
7. Duncan, J., Humphreys, G.: Visual search and stimulus similarity. Psychological Review 96(3), 433–458 (1989)
8. Enroth-Cugell, C., Robson, J.: The contrast sensitivity of retinal ganglion cells of the cat. Journal of Physiology 187(3), 517–552 (1966)
9. Frintrop, S., Jensfelt, P.: Attentional landmarks and active gaze control for visual SLAM. IEEE Transactions on Robotics 24(5), 1054–1065 (2008)
10. Itti, L.: Models of bottom-up attention and saliency. In: Itti, L., Rees, G., Tsotsos, J. (eds.) Neurobiology of Attention, pp. 576–582. Elsevier, Amsterdam (2005)
11. Jones, J., Palmer, L.: An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. Journal of Neurophysiology 58(6), 1233–1258 (1987)
12. Lagarde, M., Andry, P., Gaussier, P.: Distributed real time neural networks in interactive complex systems. In: CSTST, pp. 95–100 (2008)
13. Maass, W.: Networks of spiking neurons: the third generation of neural network models. Neural Networks 10, 1659–1671 (1997)
14. Maillard, M., Gapenne, O., Gaussier, P., Hafemeister, L.: Perception as a dynamical sensori-motor attraction basin. In: Capcarrère, M.S., Freitas, A.A., Bentley, P.J., Johnson, C.G., Timmis, J. (eds.) ECAL 2005. LNCS (LNAI), vol. 3630, pp. 37–46. Springer, Heidelberg (2005)
15. Michalke, T., Fritsch, J., Goerick, C.: Enhancing robustness of a saliency-based attention system for driver assistance. In: Gasteratos, A., Vincze, M., Tsotsos, J.K. (eds.) ICVS 2008. LNCS, vol. 5008, pp. 43–55. Springer, Heidelberg (2008)
16. Morrison, A., Mehring, C., Geisel, T., Aertsen, A., Diesmann, M.: Advancing the boundaries of high-connectivity network simulation with distributed computing. Neural Computation 17(8), 1776–1801 (2005)
17. Parikh, N., Itti, L., Weiland, J.: Saliency-based image processing for retinal prostheses. Journal of Neural Engineering 7(1) (2010)
18. Treisman, A.: Preattentive processing in vision. Computer Vision, Graphics and Image Processing 31, 156–177 (1985)
19. Vitay, J., Rougier, N., Alexandre, F.: A distributed model of spatial visual attention. In: Wermter, S., Palm, G., Elshaw, M. (eds.) Biomimetic Neural Learning for Intelligent Robots. LNCS (LNAI), vol. 3575, pp. 54–72. Springer, Heidelberg (2005)
20. Wolfe, J.: Guided Search 2.0: A revised model of visual search. Psychonomic Bulletin and Review 1(2), 202–238 (1994)
21. Yuille, A., Geiger, D.: Winner-take-all mechanisms. In: The Handbook of Brain Theory and Neural Networks, pp. 1056–1060. MIT Press, Cambridge (1998)

Fuzzy Hough Transform-Based Methods for Extraction and Measurements of Single Trees in Large-Volume 3D Terrestrial LIDAR Data

Leszek J. Chmielewski[1], Marcin Bator[1], Michał Zasada[2],
Krzysztof Stereńczak[2], and Paweł Strzeliński[3]

[1] Warsaw University of Life Sciences,
Faculty of Applied Informatics and Mathematics
leszek_chmielewski@sggw.pl
http://wzim.sggw.pl
[2] Warsaw University of Life Sciences, Faculty of Forestry
[3] Poznan University of Life Sciences, Faculty of Forestry

Abstract. This startup study suggests that more accurate and quicker methods of forestry terrestrial LIDAR data analysis can be developed, but new benchmark data sets with the ground truth data known are necessary for these methods to be validated. It follows from the literature review that the improvement in the methods can be attained by the use of newer Hough transform-based (HT) and other robust fuzzy methods for data segmentation and tree measurements. Segmentation of trees can be done by the limit fuzzification of the data around the breast height. Several HT variants having different properties can be applied to measure the diameter at breast height and the accuracies better than those offered by the commercial software seem to be attainable.

1 Introduction

Forest inventory methods require numerous and relatively accurate measurements made on trees. The parameters of the trees which are calculated from these data depend on the application. In the simplest case these are the tree height and the stem diameter at the height of human breast, that is, at 1.3 m – diameter at breast height (DBH), measured for each tree, and the number of the trees. Numerous other parameters can be of interest, for example, the height-diameter profiles, ovality of the stem, open stem height, shape and diameters of branches, diameter and other parameters of the crown, or the parameters related to the leafage. One of the measuring methods of choice is the terrestrial LIDAR (LIght Detection And Ranging) scanning [1,3,4,7,11,18,20,21,22]. Such scanning gives a 3D cloud of points indicating the surfaces of the tree stems, branches and leaves, and other objects of less or no interest possibly present in the measured area, like bushes or litter. A scan can contain millions of measured points forming a 3D image of a forest.

Probably the most frequently used technique in the analysis of the data cloud is the Hough transform (HT, the first version introduced in [8]). Its use is

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 265–274, 2010.

reported in nearly all the literature; however, it seems that even in some of the recent papers the scope of the versions of HT used is restricted, with the prevalence of the version for the detection of circles in its classical form, as introduced in 1975 [12]. Therefore, in the present paper we have tried to remind and partly check the possibilities of applying some of the more recent achievements in the domain of accumulation-based methods, for which the basic HT is a prototype (cf. [15], Chapt. 5.4). In particular, we use the fuzzy methods reported to be robust in the case of sparse measurements with a considerable content of erroneous data [6]. We present the methods operating basically around the breast height for segmenting the data into subsequent trees and measuring the DBH. The paper will be a starting point for a planned study on a larger set of measurement methods for trees, including their validation on large data sets.

2 State of the Art and New Possibilities

In some of the papers on the analysis of terrestrial LIDAR data the explicit references to the literature on Hough transform are absent [3,5,18,20], although it is reported to be used in them. The paper by Aschoff, Spiecker et al. [1] (in [19]) directly cites the book [16] where the HT for circles introduced by Kimme in 1975 [12] is described. In the paper by Simonse, Aschoff et al. [18] this basic HT for circles is explicitly reported, but without any reference to the literature on HT. This paper is cited by Bienert, Maas et al. [3] (in [13]). In the paper by Vosselman et al. [21] (in [19]) the extension of the HT by using the lines normal to the surface of the object sought, first introduced by Illingworth and Kittler for circles in [9], is used to find cylindrical surfaces. Khoshelham [11] (in [17]) extends the Generalised HT to 3D data. One of the review papers on the HT [10] is cited in this paper. Khoshelham gives attention to the problem of efficiency, so the hierarchical and probabilistic HTs are considered. Therefore, the question of the scale at which the parameter space is divided is addressed in some way. This question has also been mentioned in [21], but no solution was proposed.

The Hough transform and the derived accumulation-based methods are indeed the right choice for the application considered. The main features of the LIDAR data which support such a choice are their sparsity and the presence of gaps and errors, or noise, in the data. As reported in [6], for some versions of the HT, the simplified measure of robustness which can be the share of outliers in the data that still does not prevent the HT from yielding a correct result, exceeds 50%, and when a properly fuzzified version of the method is used it can be as high as 70-80%. The tool which can make it possible to reach such a degree of robustness is the *weak fuzzification*. Furthermore, the *strong* or the *limit fuzzification* introduced in [6] can be used to stabilise the results of segmentation of the LIDAR forestry data into single trees.

In the present paper we shall extend on the good tradition of using the Hough transform for the segmentation and for selected measurements of large-volume terrestrial LIDAR data. Starting from some preliminary results we shall highlight the deficiencies of the methods applied until now and discuss the possibilities of

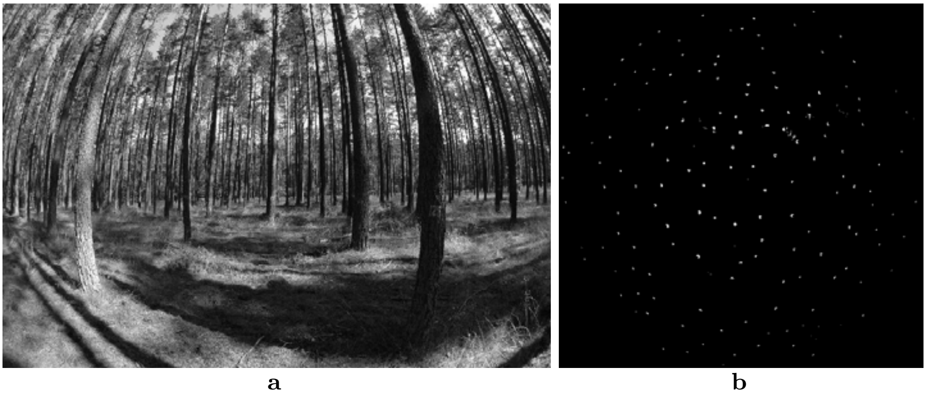


Fig. 1. (**a**) A view of a fragment of the test stand. (**b**) Section through the LIDAR data of the stand around the breast height, contrast enhanced; fragments can be seen better in Fig. 3a08 and a37.

choosing the methods still rarely used or unused in the application considered. The concepts will be illustrated with results received for sample data described in the next Section.

3 Test Data

The data were scanned at the Scots pine stand belonging to the Niedźwiady Forest District, Regional Directoriate in Szczecinek, West-Pomeranian Voivodship (North-Western Poland), with ground-based FARO LS HE880 LIDAR scanner. Nominal linear error was ±3 mm at 25 m. The data were overlaid from scans made from more than one position of the scanner. Average DBH of the trees in the stand was 26 cm and average height was 19 m. A fragment of the stand can be seen in Fig. 1a.

As a first data set, a layer of the data cloud around the breast height was cut from the whole available data. The thickness of the layer is 1 m and the breast height is 1.3 m, so the layer extends around $[0.8, 1.8]$ m above the ground level at the foot of a selected tree, further referred to as *tree37* (the ground was reasonably close to horizontal). The projection of these data onto a horizontal plane is shown in Fig. 1b. This was the *thick100* data set. Further, a 4 cm layer of the whole data around the breast height of the same tree was selected. This was the *thin004* data set.

The data on the trees in the centre are of better quality than those for the trees near the border of the region. Therefore, for further tests performed on the data on single trees, two example trees were selected (see Fig. 2c. The first tree, labelled *tree37*, chosen as an example of easy data, was near the central region and was scanned around from more than one scanner position (Fig. 3a37). The second tree, labelled *tree08*, an example of difficult data, was near the region

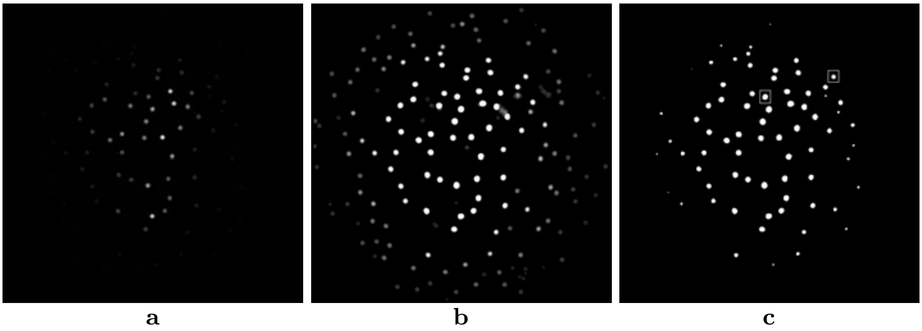


Fig. 2. (**a**) Data of Fig. 1b fuzzified up to the limit with respect to an expected thickest tree. (**b**) Image **a** with enhanced contrast. (**c**) Image **a** thresholded at 3% of its maximum giving 72 candidates for trees. Selected example trees are marked with rectangles: upper-right – difficult *tree08*, lower-left – easy *tree37*.

border and was scanned from only one side (Fig. 3a08). The data set was complemented with the results of measurements made with the software of the laser scanner and with the manual DBH measurements. However, these results were available only for the trees in the central part of the measured region, so for the *tree08* the ground truth data were unavailable.

4 Segmentation

The segmentation of the tree stand into single trees is the easiest at a height well above the ground and below the tree crowns. Such a height differs between tree species, however in the case of our data it extended from about 0.5 to 4.0 m above the mean ground level. Therefore, the breast level was a reasonable candidate height for the segmentation. A thick layer as described in Sect. 3 has been taken.

The main idea is to treat the number of measurement points projected onto the horizontal plane *Oxy* as the histogram, that is, the experimental approximation of the probability density function, of the presence of a tree surface. The trees are the largest objects in the data cloud so they should prevail in the histogram. In fact, in the global histogram the local histograms for single trees are nearly separated. Let us concentrate on a single tree (yet not localised). To find the location of the tree stem, not its surface, the histogram can be fuzzified with the fuzzifying function in the shape of the positive part of an inverted paraboloid, with the support half-diameter not smaller than the the maximum expected tree diameter. Then, it is guaranteed that a tree manifests itself in the fuzzified histogram as a single paraboloid because this process is the *limit fuzzification* of the histogram [6] with respect to a single tree. Consequently, to find all the trees it is enough to fuzzify the image like the one in Fig. 1b up to the limit with respect to an expected thickest tree, and to find maxima in the resulting image like that in Fig. 2a.

Each significant maximum corresponds to one tree. A simple thresholding can give a satisfactory segmentation into regions belonging to separate trees. In more difficult cases, the watershed method could be applied. The points which project themselves onto the region belonging to one maximum belong to one tree. The histogram does not have to be of large resolution: several centimetres per pixel are enough.

5 Verticalization

The trees are not exactly vertical. This makes the projections of the data points belonging to a layer of 1 m thickness lie far from the expected nearly-circular shape. This was not detrimental in the process of segmentation, while in the calculation of the diameter the inclination of the trees should be compensated for, if a thick layer of the data is to be used. In the case of the trees having the data points from many sides, as *tree37*, a thin layer can be used. However, for the trees with less data points, like *tree08*, a thick layer can be necessary.

The data points segmented out for a single tree were projected on the vertical planes and fuzzified. The dominating straight lines were found. The angles of these lines were used to recalculate the data points to the coordinate system parallel to the stem axis thus found. The centre of rotation was at the mid-height of the data layer.

6 Measurements of the Diameter at Breast Height

To determine a circle it is enough to know its three points to form three equations for three parameters of a circle. Three points (three pixels) are called the *elemental subset* for a circle [14]. If an elemental subset is used to vote for a geometrical figure in the HT, then it can vote for a full set of parameters of this figure. In general, the subsets of the cardinality equal or smaller (never larger) than that of the elemental subsets are used as voting subsets in HT. The smaller the cardinality of the voting subset, the more complex the geometrical figure is plotted in the parameter space. Not all the parameters of a figure sought must be found at once. For example, if only the centre of a circle is of interest, then the elemental subset can be just two points, but complemented with the directions of the normals to the circle in them.

In the cited literature on LIDAR measurements of trees it is always assumed that a cross-section of a tree is circular. This is not the case in general. When a circular tree is not vertical, its horizontal section is an ellipse. Further, a simple observation indicates that trees are not regular objects (see Fig. 3b08 and b37. The next more complex approximation of a horizontal tree section, after a circle, is an ellipse. In [15], Chapt. 5.4.3-5.4.5, a number of versions of the HT for circles and ellipses have been described. Some of their basic features are compared in Tables 1 and 2.

It has been taken into account that the centre of the figure sought is of primary interest and the other parameters are easy to find in the further, simpler accumulation steps.

Table 1. Comparison of the features of HT for circles. PS: parameter space. Expectations on the robustness inferred merely from the features of the method.

No.	Voting subset	Figure in PS	Remarks on the data	Exp. robustness
1	1 point	cone	raw points can be used	high
2	1 normal	line	projected on a plane, fuzzified	moderate
3	2 points	line	raw points can be used	low
4	2 normals	point	projected on a plane, fuzzified	low/moderate

Table 2. Comparison of the features of HT for ellipses

No.	Voting subset	Figure in PS	Remarks on the data
1	2 points with their normals	line	projected on a plane, fuzzified
2	4 points	line	raw data can be used

The comparisons implies several conclusions. Bearing in mind the quality of the data it can not be expected that all the points in a pair or a quadruple appear to lie very near to the circle of ellipse to be found, or that they lie at the same side of its border. Therefore, the stability of the result of finding the figure plotted in the parameter space can be low, adversely influencing the robustness of the whole method. Finding the normal to the figure border can be difficult unless the the data are fuzzified, which makes it possible to find the normal as the gradient of the intensity of a resulting image. Fuzzification necessitates for projecting all the points onto one plane. This process reduces the number of data and the resulting processing time, but with the simultaneous loss of information. Decrease of the quality of the voting process in the parameter space must be compensated with the *weak fuzzification* [6]. Finally, the more points in one voting set, the larger the number of all possible sets and hence the longer the processing time. The processing time increases also with the complexity of the figure plotted in the parameter space.

In the present preparatory study the methods 1 and 2 from Table 1 have been used. These are the method of Kimme [12] and Illingworth and Kittler [9], respectively. The other methods will be tested in the next stages of the work.

7 Results and Discussion

For the segmentation, images with resolution of 10 cm/pixel were formed from the *thick* data set. For the limit fuzzification with respect to a single tree, the fuzzification function in the form of an inverted parabola, as advised in [6], was used, with the support of 11 pixels, that is, 110 cm. No tree was expected to have a larger diameter, so one single maximum was detected for each tree. The fuzzified image was thresholded at 3% of the maximum for these data. In this way, 72 tree candidates were found, as shown in Fig. 2.

Table 3. Results for *tree37*. Besides the results calculated with the methods from Tab. 1, No. 1 and 2, the manually measured result (ground truth) and result calculated with the software of the LIDAR (L. soft) were available.

method	Manual	L. soft	HT cone		HT line	
data set			*thin004*	*thick100*	*thin004*	*thick100*
result [mm]	238	221	248	248	232	232
error [%]	0.0	-7.1	4.2	4.2	-2.5	-2.5
time [s]			68	1552	8	8

Table 4. Results for *tree08*. Only the results calculated with the methods from Tab. 1, No. 1 and 2, were available; the ground truth was unknown.

method	HT cone		HT line	
data set	*thin004*	*thick100*	*thin004*	*thick100*
result [mm]	356	364	332	360
time [s]	17	372	8	14

For the verticalization of the *thick100* data set, the image resolution was 2 mm/pixel. The angle resolution was 1°. The paraboloidal fuzzification function with the support of 31 pixels, that is 6.2 cm was used. Inclination angles of the stem appeared to be up to ±15°.

For the diameter measurements the resolution and the fuzzification function were the same as for the verticalization. The measurement results are shown in Tables 3 and 4. The *thick100* data were verticalized, while the *thin004* ones were obviously not. Times of calculations are given for an Intel®Core™2 Duo Pentium at 2.4 GHz.

In general, the results obtained should be validated on a large set of ground-truth test data, with the properly formulated measure of correctness of the results (discussed for example in [2]). The results shown here are preliminary and can be treated only as an encouraging feasibility study.

The segmentation results seem very satisfactory at this stage. Simple thresholding of the limit fuzzification results should be a sufficient method in most cases, while the watershed segmentation can be used in case of necessity.

Among the verticalization results, from 8 to 11% of angle measurements were erroneous (7 to 10 in 90). It is expected that better results could be attained if the images with only the vertical edges of the tree were used instead of the the raw images, fuzzified.

The breast diameter measurements of the *tree37* with the methods presented here seemed to be slightly better than those received with the software package provided together with the LIDAR, with the HT with lines performing the best. For the more difficult *tree08* the ground truth data were not available, so it can only be stated that the result for the HT with lines for the data set *thin004*, having much less measuring points, is far from the cluster of the other results. This indicates than this method can be less robust in difficult cases.

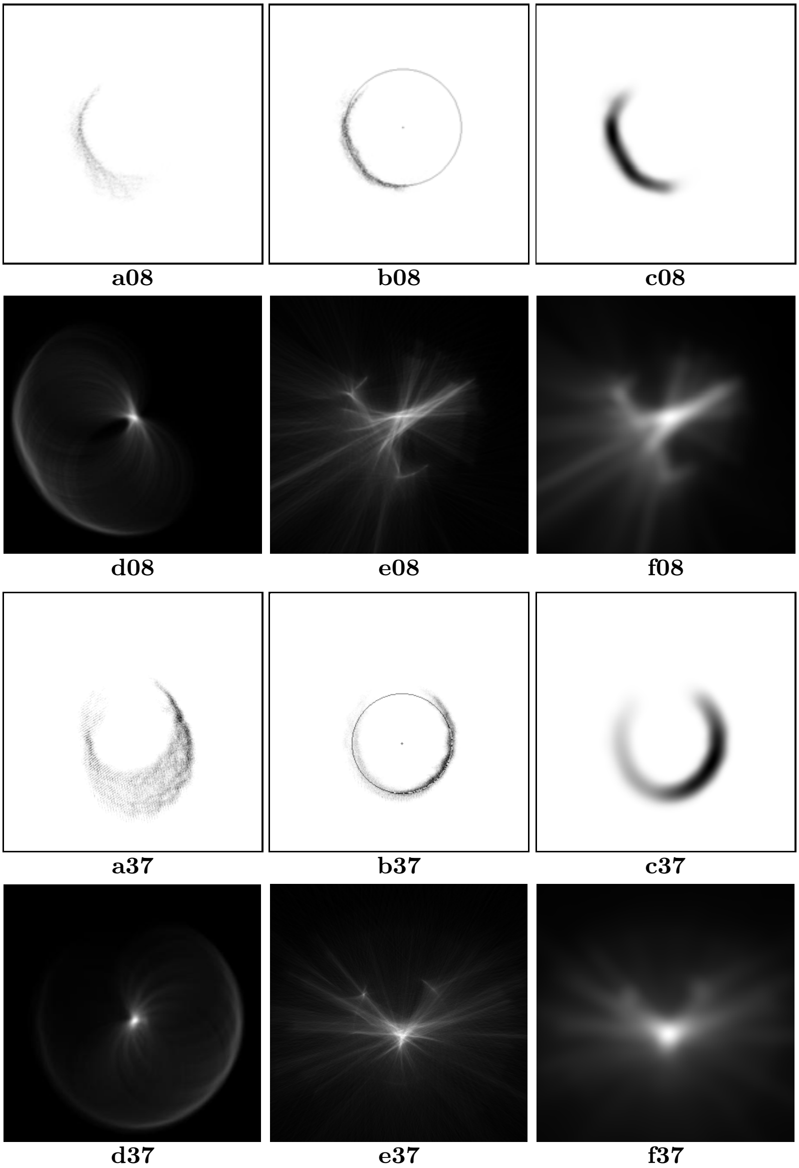


Fig. 3. Calculations for *tree08* and *tree37*, data set *thick100*. (**a**) Raw data points projected on a horizontal plane *Oxy*; (**b**) Data points of image **a** after verticalization; result calculated from the accumulator **f** is overlaid. (**c**) Fuzzified data of image **b**. (**d**) Accumulator of the HT with circles calculated directly from the data **a** – layer containing the maximum; fuzzification not necessary. (**e**) Accumulator of the HT with normals calculated from fuzzified data **b**; fuzzification is needed due to the presence of many maxima. (**f**) Accumulator **e** fuzzified.

In general, the 2 mm accuracy in DBH seem to be more than enough for such irregular objects like trees, where the actual cross-section is rarely a regular circle. However, such accuracy is usually treated as a standard in forestry. The present errors of the methods transcend this standard, so the image resolution of 2 mm seems to be enough in the considered application.

The calculation times for the HT with cones can be reduced if the data are projected on the horizontal plane, as in the HT with lines. This could have little or no influence on the accuracy if the image resolution is sufficiently high.

Trials with more variants of the Hough transform are necessary, including the use of an ellipse as a model of a tree section (see Tabs. 1 and 2).

8 Conclusions

Tree measurements can be speeded up with the use of terrestrial LIDAR-based measurements, but the accuracies attainable with some currently available software seems to be questionable. The presented startup study indicates that the segmentation of the terrestrial LIDAR data into separate trees is not a difficult task, so that the starting points for the analysis of single trees going up from the breast height can be easily found in the cloud of the measuring points. The literature study suggests that the application of more advanced Hough transforms for circle and ellipse detection can lead to better results and quicker calculations of the breast-height diameters.

Using the more advanced methods, including the fuzzy Hough transform-based and other robust techniques seems to be promising, but more work should be done. The impact should be laid upon the development of large benchmark data sets with the known ground truth and on the validation of the existing and new methods against the credible data.

References

1. Aschoff, T., Spiecker, H.: Algorithms for the automatic detection of trees in laser-scanner data. In: Thies et al [19], pp. 71–75
2. Bator, M., Chmielewski, L.J.: Finding regions of interest for cancerous masses enhanced by elimination of linear structures and considerations on detection correctness measures in mammography. Pattern Analysis & Applications 12(4), 377–390 (2009), doi:10.1007/s10044-008-0134-x
3. Bienert, A., Maas, H.-G., Scheller, S.: Analysis of the information content of terrestrial laserscanner point clouds for the automatic determination of forest inventory parameters. In: Koukal and Schneider [13], pp. 55–60
4. Bienert, A., Scheller, S., et al.: Application of terrestrial laser scanners for the determination of forest inventory parameters. In: Maas, H.G., Schneider, D. (eds.) Proc. ISPRS Commission V Symposium on Image Engineering and Vision Metrology, pp. 25–27 (September 2006)
5. Bienert, A., Scheller, S., et al.: Tree detection and diameter estimations by analysis of forest terrestrial laserscanner point clouds. In: Rönnholm et al. [17], pp. 50–55

6. Chmielewski, L.J.: Fuzzy histograms, weak fuzzification and accumulation of periodic quantities. Application in two accumulation-based image processing methods. Pattern Analysis & Applications 9(2-3), 189–210 (2006) doi:10.1007/s10044-006-0037-7
7. Danson, F.M., Hetherington, D., et al.: Three-dimensional forest cannopy structure from terrestrial laser scanning. In: Koukal and Schneider [13], pp. 61–65
8. Hough, P.V.C.: Machine analysis of bubble chamber pictures. In: Proc. Int. Conf. on High Energy Accelerators and Instrumentation. CERN (1959)
9. Illingworth, J., Kittler, J.: The adaptive Hough transform. IEEE Trans. PAMI 9(5), 690–697 (1987)
10. Illingworth, J., Kittler, J.: A survey of the Hough transform. Comp. Vision, Graph., and Image Proc. 44(1), 87–116 (1988)
11. Khoshelham, K.: Extending Generalized Hough Transform to detect 3D objects in laser range data. In: Rönnholm et al. [17], pp. 206–210
12. Kimme, C., Ballard, D., Sklansky, J.: Finding circles by an array of accumulators. Comm. Assoc. of Computing Machinery 18(2), 120–122 (1975)
13. Koukal, T., Schneider, W. (eds.): Proc. Int. Workshop on 3D Remote Sensing in Forestry, Vienna, February 14-15. EARSeL SIG Forestry and ISPRS WG VIII/11 (2006)
14. Meer, P.: Robust techniques for computer vision. In: Medioni, G., Kang, S.B. (eds.) Emerging Topics in Computer Vision, pp. 107–190. Prentice Hall, Englewood Cliffs (2004)
15. Nixon, M., Aguado, A.: Feature Extraction & Image Processing. Newnes, Oxford (2002)
16. Pitas, I.: Digital Image Processing Algorithms and Applications. John Wiley & Sons, Inc., New York (2000)
17. Rönnholm, P., Hyyppä, H., Hyyppä, J. (eds.): Proc. ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007, part 3/W52 of IAPRS, Espoo, Finland, September 12-14, vol. XXXVI (2007)
18. Simonse, M., Aschoff, T., et al.: Automatic determination of forest inventory parameters using terrestrial laserscanning. In: Proc. ScandLaser Scientific Workshop on Airborne Laser Scanning of Forests, Umeå, Sweden, pp. 251–257 (2003)
19. Thies, M., Koch, B., et al. (eds.): Laser Scanners for Forest and Landscape Assessment. In: Proc. of the ISPRS working group VIII/2, part 8/W2 of International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Freiburg, Germany, October 3-6, vol. XXXXVI (2004)
20. Thies, M., Spiecker, H.: Evaluation and future prospects of terrestrial laser scanning for standardized forest inventories. In: Thies et al. [19], pp. 192–197
21. Vosselman, G., Gorte, B.G.H., Sithole, G., Rabbani, T.: Recognising structure in laser scanner point clouds. In: Thies et al. [19], pp. 33–38
22. Zawiła-Niedźwiecki, T., Stereńczak, K., Bałazy, R., Wencel, A., Strzeliński, P., Zasada, M.: The use of terrestrial and airborne LIDAR technology in forest inventory. Ambiência 4, 57–68 (2008)

Image Recognition Techniques Applied to Automated and Objective QoE Assessment of Mobile WWW Services

Michał Choraś[1], Adam Flizikowski[1], Rafał Kozik[1], and Witold Hołubowicz[2,3]

[1] Institute of Telecommunications, UT&LS Bydgoszcz
chorasm@utp.edu.pl
[2] Mickiewicz University, Poznań
holubowicz@amu.edu.pl
[3] ITTI Ltd., ul. Rubież 46, 61-612 Poznań, Poland

Abstract. In this paper the innovative methodology based on image/video processing techniques for objective measurements of QoE (Quality of Experience) dedicated for mobile terminals is presented. The proposed methodology has been used to evaluate WWW services offered by mobile operators in Poland. The clear advantages of applying image processing techniques to such a task are objectiveness of the methodology and independence of the underlying network infrastructure.

1 Rationale

The mobile network technologies are continuously evolving towards high bandwidth, QoS enabled, mobility rich 4G solutions that will revolutionize capabilities of novel services in wireless networks. In parallel the end-terminals portfolio has exploded with devices providing greater functionality and usability with more processing power on board. Panoramic displays touch screens and TV-out ports are steps towards providing improved user satisfaction.

Measuring and providing quality of service to customers of network operators and service providers is a crucial aspect of current telecommunication market. However, the real user experience is not only driven by combined performance of underlying access network and terminal characteristics. The complexity of internal processing of a web browser used to render data to a user may influence the quality of experience (QoE). Thus to assess the real user experience of the HTTP services it is not enough to analyze packet level metrics on the mobile terminal (or inside network) as implementation of various browsers may affect efficiency of data rendering differently.

There is shortage of methods for assessing subjective (as perceived by end user) quality of data services like WWW. In addition capabilities of automated measurement are limited due SW/HW characteristics inherent to modern mobile terminals. Among all the key limitations are: insufficient processing power of mobile phones/devices, limited multi-threading support etc.

On the other hand European directives and national regulators (e.g. see [1] for polish regulations) impose the obligation to ensure the quality and regular

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 275–282, 2010.

publication of the quality indicators of the service portfolio offered by operators of telecommunication networks and service providers. With the growing number of services available in the network the complexity of measuring and control their costs, ultimately incurred by subscribers, and the indicators of quality of service, collected by different entities have different forms of presentation and various reliability for observers from the outside.

Authors propose the novel idea of applying concept of image processing to the analysis of visual information as originally seen by the end user on the mobile terminal screen. The paper presents the concept, system design, verification and validation and preliminary results of the proposed techniques. The presented results show that the use of image processing techniques has numerous advantages against complementary user assisted measurement techniques for objective QoE measurement.

The paper is structured as follows: in section 2 motivation for our novel image-based Quality of Experience (QoE) assessment for mobile devices is clearly described. In Section 3 our image based methodology is presented. In Section 4 practical deployment (MEVA tool) is shown. Evaluation and conclusions are given thereafter.

2 Motivation for Image Recognition-Based QoE for WWW

According to the SEOptimize web portal, both proprietary and open source desktop software will witness a further rise in popularity as browser based web applications and SaaS (software as a service) will become more common ([2]). The browser will become a second OS with both web based applications and software being added as browser extensions. Thus sample trend analysis shows that it is essential to evaluate quality of web-based services exactly as perceived by end user.

The latter, however, is not possible without actually assessing broad aspects of web browser application architecture. It is important to understand that focusing only on IP level tracing (data packets that the mobile browser message exchanges with external servers) limits the reliability of the achieved results. The total page download time depends both on the network (access, content) performance as well as the browser application architecture.

Currently the user experience is mainly determined by looking at the Quality of Service of the network (ITU G.107 e-Model, PESQ, PSQM). A portfolio of commercial tools exists that offers automated assessment of QoS based on the above metrics (TEMS Investigation, QVoice etc.).

However there is misleading hypothesis that QoE will increase with increasing QoS, while this is not necessarily true: "It is possible to have excellent QoS and poor QoE" [3]. On the other hand QoE is defined as "general service application performance" thus it is defined by service accessibility, availability and integrity [4]. Although even access to such parameters from the monitoring probes or system in the network or end terminal during service consumption can be not enough extension towards improved user experience without reliable QoE assessment techniques. Not only network performance related metrics are correlated to the user experience, in [6] the user experience is linked to the video quality within a video conferencing system. From a social research point of view, user experience is often linked to the usability of the application [6] and the way applications and technology merge and assist in every day live. Complementary to technical research, the user context (age, profile, activities, culture) is relevant to the experience [7].

Moreover, the structure of browser functional components indicates that focusing only on measurements of the network aspects does not enable reliable user experience analysis (see the indications of the QoE/QoS capabilities respectively to the probe placement). Not encompassing the analysis of the process of web page rendering in the page download time measurements may lead to unreliable QoE results.

The above analysis demonstrates that on one hand status of current mobile web browsers varies (from heavy-footprint and non-fully standard compliant built-ins to highly optimized, transcoding-aware mobile browsers). However, it is important that due to layered nature of data processing inside browser application and difference in optimizations level of page rendering process, the quality of HTTP service observed by end user may be affected.

Therefore hereby, we propose innovative approach to QoE assessment by applying image-based methodology as presented in the next sections.

3 Image-Based Methodology for QoE

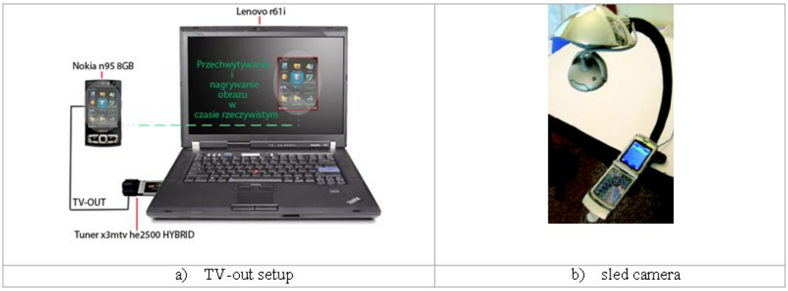
The major contribution and focus of this paper is the application of image/video processing techniques to real-life industrial problem. In our objective and automated QoE methodology, image processing techniques are used at the preprocessing (detection) and post-processing stage to calculate features.

The following objective QoE measures have been applied:

- download time measurement (based on the ”progress bar” analysis)
- throughput measurement (based on ”progress bar” analysis)
- service availability measurement

Term ”progress bar” refers to the graphical element in HTTP browser that indicates progress of downloading web pages. We propose to perform objective tests using support of image recognition methods in order to gain precision, reliability and repeatability of the statistics calculation. The following two video capture configuration options can be envisaged for performing the measurements properly (Fig. 1). It is possible to perform video capture using both TV-out (Fig. 1a) and a sled camera mounted over the phone (Fig. 1b). However, the option with camera is more demanding with respect to its usage during field testing. Suitable environment conditions as well as increased programming effort are required to get reliability of image recognition comparable to the option ”a”.

Authors decided to focus on validating the TV-out solution as it provides highest quality, is not sensitive to environmental conditions (light reflections,



a) TV-out setup

b) sled camera

Fig. 1. Video capture configurations

movements) and does not require special treatment (handle, camera movement mitigation, protection from excessive light, way of doing tests) from the test operator. The input video stream captured using TV-out has size of 720x576 pixels (PAL format), but as browser area is about 300x450 pixels significant reduction of processing data is possible (about 60% reduction of required memory and processor resources). The respective mobile terminal selected for assembly of the prototype configuration was the Nokia N95 8GB. Objective tests using the above method were carried by test engineers who performed sequence of repetitive actions according to the testing methodology.

The measurement process included the following phases:

- data acquisition - process of opening web pages with each of four mobile terminals (one terminal per each network operator) is recorded using the input video stream captured directly from TV-out interface of the phone. This functionality enabled storing video sequence produced by terminal operator (test engineer),
- post processing - after the test cases at various locations are completed image normalization and post-processing of the stored video captures is performed,
- image recognition - in order to detect web page downloading completion status (content downloaded and rendered within the browser) there are two complementary ways by proceeding in an semi-automated way.

Download time measurement, as a key metrics, was captured using technique of tracking the progress bar. It is assumed that web browser's "progress bar" is a reliable source of information about the completeness of the web page download status (in comparison to partial information provided by sniffing HTTP conversation at the packet level). Thus the moment of 100% completed progress bar is treated as the web page download completion. The method assumes that video frames of download process will be tracked (Fig. 2a,d) in order to identify the very moment when progress bar is completed (feature based: e.g. color). During the verification tests it was proven that there is no need to consider the alternative method based on differential video frames as the progress bar is suitable to perform proper measurement during all test cases.

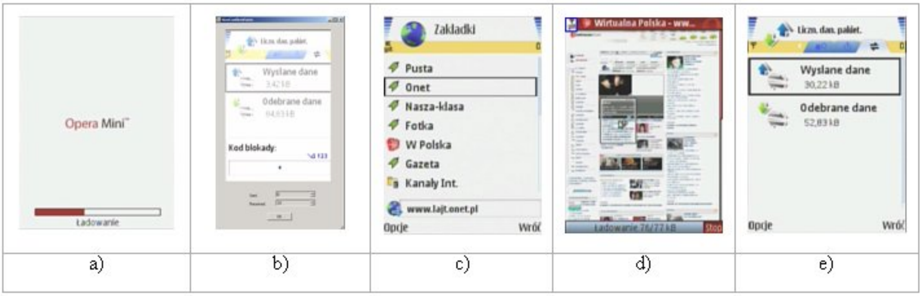


Fig. 2. Sample splash screen (a), OCR control box (b), bookmarks (c), sample page (d) and counters (e)

4 Practical Deployment: Mobile Equipment Video Analysis (MEVA) Tool

For validating the concept of image recognition oriented QoE measurement, authors have developed prototype software tool - the Mobile Equipment Video Analysis tool (MEVA). The MEVA application implements video processing functionalities that are applied on the video streams captured during field tests. After field tests are finished and results stored as video files it is then opened within the MEVA application context in batch mode and post processing starts. The key functionalities and supported image processing operations of MEVA are described below:

- Browser recognition. It is the first important step using advanced video processing. This step is executed at the beginning of every sequence and targets recognition which of available (defined in the system) browsers is currently in use. MEVA is able to recognize three the most popular (in Poland) mobile browsers: Opera Mini, Blue Connect and Nokia build-in browser. Each of browsers use different color theme, also splash screen or menu structure are unique for every browser. It is essential to differentiate browser type because page loading process looks completely different in various browsers (order of performed operations may be different, some stages are specific for single browsers only etc.). MEVA captures splash screen and then compares it with defined patterns.
- Network technology recognition. Information about technology used during the packet switched data session (GSM, UMTS or HSDPA) is displayed on the screen of the mobile phone in the left top corner. Technology info has a form of a small icon - MEVA contains a set of pattern definitions of these icons and the recognition process relies on the comparison of the icon from the screen with mentioned set of patterns. If the icon from the screen matches any pattern, then MEVA will store this information, and this stage is successfully finished. If process for current video frame fails (e.g. the icon was too blurred and MEVA was not able to recognize it correctly) process will

be re-executed for the next one (single session usually consists of hundreds of frames so there is only hypothetic chance to fail).

- Page name recognition. Technique used here strictly depends on the particular browser being captured. For some browsers (Opera, BlueConnect) this process uses pattern recognition (similar to technology recognition), but for Nokia Build-in browser OCR (Optical Character Recognition). Pattern recognition: when page is loaded, page name (title) can be found on the top bar of the browser as well as so called "favicon" (small icon with page's logo). MEVA is able to find such favicons, and then recognize it by comparing with the build-in set of known favicons. Process mostly relies on color histogram analysis (technology recognition stage is based on shape recognition). Logos of different web pages have different predominant color schemas (e. g. polish portal Interia has yellow as predominant color, another one WP is red and so on). MEVA simply analyzes frequency/intensity of each color on the favicon area (in percents) and then using some set of decision rules is able to find respective page name.
- Page size recognition. MEVA ability to recognize page size allows comparing single page's size in different browsers, for different technologies and operators. What is even more important this feature allows calculating connection speed. By dividing page size and loading time network bandwidth can be calculated. To recognize size of the currently loaded page MEVA waits till operator closes the browser and goes through telephone's menu to the valid widget displaying transferred data amount. MEVA is able to automatically recognize this page's content (in all Nokia phones this menu looks similar).
- Loading time measurement. At this stage time needed to load web page is calculated. Loading time measurement is one of the key functionalities provided by MEVA. This step allows gathering information about page loading time in order to compare results between different browsers and operators. MEVA distinguishes two types of time variables for page download: connecting time and loading time. Connecting time is measured from the moment, when empty progress bar appears on the screen and continues to the moment, when some progress will be notified on it (color inside of progress bar will be changed). The second phase: loading time is measured from the moment, when connecting phase ends to the moment when progress bar disappears from the screen.

Especially page size recognition was a challenge because the OCR library used (Microsoft MODI library) produced results far from perfect due to video quality deterioration of signal captured from TV-out interface (analog) and the phone menu characteristics, background themes and colors as well as the font sizes (especially the dot sign position in the recognized number) - Fig. 2e. In order to improve the reliability of page size recognition authors have on one hand implemented alternative manual confirmation mechanism of recognized text/numbers (message box is presented in the Fig. 2b), on the other a set of hardcoded rules were specified (e.g. "if page X has size Y at the calibration stage then with high probability it cannot change by tenfold in subsequent openings").

5 Evaluation

The purpose of preliminary measurements was to assess the satisfaction of a typical WWW service user in a mobile phone. It focused on the end customer perspective. The selected parameters of WWW service were measured in four cellular networks of polish operators. During the test campaign authors have verified proposed testing methodology in seven regions in Poland. More than twelve thousands observations were collected in total using the four test terminals with data service subscriptions of each polish cellular provider. Each and every measurement observation was captured and aggregated in the form of video sequences stored on a number of DVD disks. All videos from the test campaign were processed using MEVA tool. The tool has proven to provide reliable QoE measures in semi-automated fashion. The MEVA-tool operator needed to continuously monitor and assess the recognition process (page sizes and the validity of the video sequence). The maximum feasible ratio of video capture running time to its processing time is 5:1 which means that a video sequence of 25 minutes long can be processed in 5 minutes. However such optimization of the post-processing stage is not recommended due to limited error correction capabilities required from the operator (as the video frames are hard to follow while recognition efficiency is less than 100%). Authors suggest using maximum of 3-4 speedup factor (MEVA setting) for proper video processing. With such a speed operator can effectively monitor and correct result of recognition if needed. Although time needed to process large amount of videos is significant (with test engineer corrections from time to time).

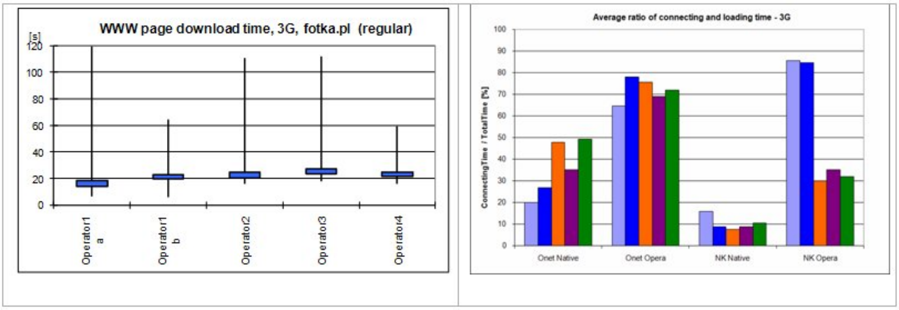


Fig. 3. Sample results: (a) download time, (b) connecting/loading time ratio

Sample results of the tests are presented in the Fig. 3. The use of proposed method of image based calculation of the key WWW performance indicators had improved the measurement fidelity and measurement error was limited to the extent dependent mainly on the level of occurrence of the image processing artifacts. In the box plot (Fig. 3a) the results of average loading time were displayed against different web browsers and mobile network operators. The left part of the chart consists of results for native browsers (higher loading time due to limited compression) whereas right most part of the data points shows loading time with Opera Mini browser (lower loading time due to server side compression). The next figure (3b) shows that loading time includes another substrate namely connecting time. As can be seen the ratio of connecting time (time needed to establish logical connection) to the loading time is not constant but changes depending on the page size and browser type used. Thus this ratio should be accounted for in the process of calculating the effective loading time (without the connecting component which mostly lowers the effective throughput).

6 Conclusions

The applicability of the novel approach/method for image/video-based QoE assessment of the mobile web browsers has been positively verified and validated during the comprehensive test campaign. Capturing video stream of the mobile's display has enabled performing WWW service tests while preserving real end-user perspective (exactly as observed by end user on the browser's display) without qualitative loss on measurement information observed with QoS only solutions (e.g. using TEMS to measure HTTP performance on packet level).

Moreover, significant improvement in automation of processing the large data set of collected observations has been achieved. The MEVA tool has further potential for introducing enhancements towards anomaly detection and real-time support of the measurement process (feedback to the test engineer in case of incomplete/failed sequence). Additional benefit of using proposed methodology is the capability of storing 100% complete, perfectly visual (easily verifiable) data from measurements in a form of video sequences that can later be re-analyzed.

Our innovative image/video processing methodology for QoE measurement and the MEVA tool have been used in practice by the telecommunication operator to evaluate its HTTP/WWW services for mobile devices.

References

1. Polish Telecommunication Law, Act from the 24th of April 2009
2. http://www.seoptimise.com/
3. Empirix, Assuring QoE on Next Generation Networks -White Paper, [online] (2003)
4. Soldani, D., Li, M., Cuny, R.: Qos and Qoe Management in Umts Cellular Systems. John Wiley & Sons, Chichester (2006)
5. Watson, A., Sasse, M.: Measuring perceived quality of speech and video in multimedia conferencing applications. In: Proc. of 6th ACM International Conference on Multimedia, Bristol (1998)
6. Monk, A., Carroll, J., Parker, S., Blythe, M.: Why are mobile phones annoying? Behavior and Information Technology (2004)
7. De Moor, K., De Marez, L.: Quality of (user) Experience: conceptualization and measurement of rising concept. In: Proc. of IAMCR Conference, Paris (2007)

Detection of Near-Regular Object Configurations by Elastic Graph Search

Gerardus Croonen* and Csaba Beleznai

Austrian Institute of Technology - AIT, Vienna, Austria
gerardus.croonen@ait.ac.at

Abstract. In this paper a novel non-rigid matching technique is presented to reliably detect near-regular object configurations in images in the presence of substantial clutter. The objects in near-regular configurations span a grid with not necessarily congruent grid cells. In addition, the output of object detectors is typically associated with ambiguities, causing occasional spurious false alarms. We propose a detection approach of spatially extended near-regular structures which reliably deals with missing objects and clutter. The search for the optimal local spatial object arrangement is formulated as a non-rigid matching problem and it is guided by the criterion of structural energy minimization while maximizing image likelihood. The applicability of the approach is demonstrated by presenting experimental results for the detection of near-regular object configurations.

1 Introduction

In this paper we address the problem of recovering the exact geometry of multiple objects located on a near-regular geometrical configuration by determining a non-rigid match between a model graph and observed object locations using the constraint of near-regularity of the objects and their salient features. Regularity (or near-regularity) is a strong cue which greatly reduces the size of the search space. We present a way to represent near-regular structures (some examples are shown in Figure 1) in a model graph and provide an algorithm that searches for occurrences of the model graph in the output of an object detector. The algorithm is robust to the variations in object position in near-regular object configurations and to noise. This paper is structured as follows. First, we provide an overview of related work. Next, we describe the proposed matching algorithm and finally we show experimental results for the detection of near-regular object configurations.

1.1 Related Work

Our approach is related to deformable template matching. The concept of deformable template matching has received much attention in the computer vision

* This work was supported by the FFG COMET project ECV.

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 283–291, 2010.

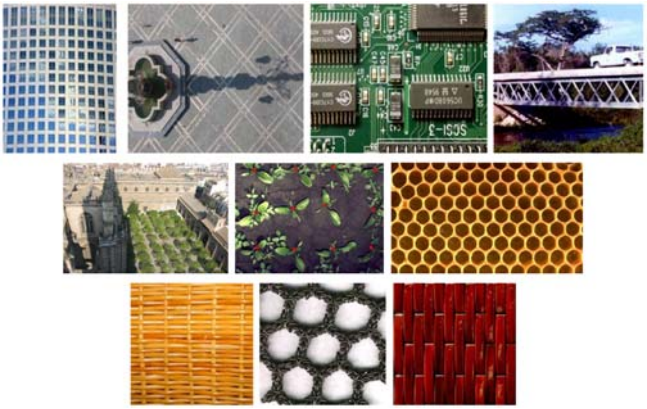


Fig. 1. Some examples of near regular object configurations

community since representing variable shape and appearance is a key functionality to achieve object detection and recognition across view angle and shape variations. Since Fischler, Elschlager [1] and Widrow [2] proposed the concept to represent parts of an object by a deformable configuration, the subject has evolved into a large research domain. Various methods have been developed to find the optimal match between an object model and a subset of observed salient points in images in an efficient manner, such as dynamic programming [3], [4] and belief propagation [5]. Numerous extensions of the original concept have been proposed in form of the Active Shape Model [6] and elastic bunch graph matching [7]. Recently, part-based visual object recognition has adopted the unified concept of using object appearance and deformable object geometry as is the case in Partial Surface Models (PSM) [8]. Our approach is similar to deformable template matching in that it applies an energy function in finding the non-rigid mapping between image features and a (grid) model graph. Within the context of deformable template matching each potential transformation of the grid model is associated with an energy measure by means of an energy function. The objective of matching is to recover the transformation with minimum energy among the set of all possible transformations. This transformation is the most likely hypothesis of object cofigurations. This work is different from related work in that its focus is on the detection of object configurations under challenging circumstances by searching for near-regularity. Related work is mainly aiming at improving object detection using combined appearance and geometrical object representations that can incorporate the large appearance variations caused by changes in view angle, scale and lighting conditions.

2 Methodology

Given an N-dimensional space in which a scalar evidence function Ψ (such as probability or similarity measure) for the occurrence of an object is defined.

The value for $\Psi(\bar{p})$ is the evidence for the occurrence of an object at the position vector $\bar{p} = [i_1 \ldots i_N]$. The positions of the objects in this N-dimensional space are nearly regularly distributed and the parameters of these distributions (their position mean and deviation) are known. Our method is applied under the assumption that the local maxima in Ψ correspond to true object positions. Near-regularity prior is represented by a graph, where nodes correspond to object positions. The objective of the algorithm is to find the subset of local maxima in Ψ that best matches the geometry of the model graph among the set of hypotheses that can be generated from the local probability maxima.

2.1 The Model Graph

The nodes of the model graph correspond to the mean object positions. The edges represent the penalties imposed for changing the distance between their associated nodes and thereby constrain the deformation of the model graph. The algorithm can be used with model graphs of any number of nodes and any combination of edges depending on the requirements of the application. Some examples of model graphs are displayed in Figure 2.

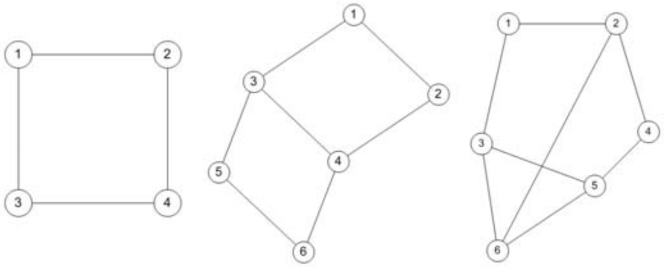


Fig. 2. Some examples of model graphs. The nodes of the graph correspond to object locations while edges represent spring-like structural constraints.

2.2 Generation of Hypothesis Graphs

Hypothesis graphs are generated from the local maxima of Ψ using the model graph. This is done by first determining the positions and values of the local maxima of Ψ. The model graph is then consecutively placed with every one of its n nodes at every one of these m local maxima. This results in nm model graph *placements*. Now for every placement of the model graph, the Euclidean distance between every model graph node and every local maximum is calculated. The local maxima of Ψ that are found to be within a maximum distance of a node are called *inliers* to this node. A list of such inliers is stored for every model graph node. If no inliers were found for a particular node, the attributed position of the corresponding node in the model graph is taken as the only inlier position. Finally, from the lists of inliers for the nodes of the model graph all the hypotheses are exhaustively enumerated.

2.3 Hypothesis Graph Energy

Every hypothesis graph can be thought of as a deformed version of the model graph. Given the model graph and a hypothesis graph, the amount of similarity between the two is defined using an energy based function on the matching edges of both graphs. The energy function needs to allow deformations of the model graph, while at the same time taking into account the local evidence in Ψ for the occurrence of an object. By definition the hypothesis graph with the lowest energy is the best match to the model graph. To satisfy these requirements, the energy function consists of two terms. The first term penalizes deformations of the hypothesis graph from the model graph, while the second term rewards hypothesis graphs that explain the observed local probability maxima. The term that penalizes deformations is based on edge length differences between matching edges (See Figure 3). An edge in the hypothesis graph that has a length equal to the corresponding edge in the model graph should have an energy equal to zero. A suitable measure between two edges satisfies these properties:

$$E_{edge}(i,j) = K[|\bar{p}_i - \bar{p}_j| - |\bar{q}_i - \bar{q}_j|]^2, \tag{1}$$

in which K is the energy penalty weight for a specific edge, $\bar{p}_n$ is position attributed to the n-th node of the hypothesis; $\bar{q}_m$ is the position attributed to the m-th nodes of the model graph. i and j are the node indices between which the edges are defined. This measure is comparable to the mechanical energy stored in a spring and is therefore referred to as the edge energy.

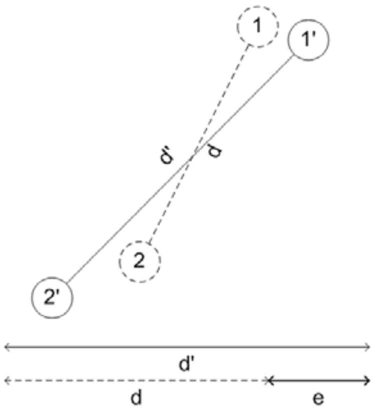


Fig. 3. Model edge (solid) and corresponding hypothesis edge (left)

The internal energy of a hypothesis graph is a measure for the geometrical correspondence to the model graph. This measure is defined as the total energy required to deform all the vertices of the hypothesis graph to obtain the model graph. This energy measure is defined as the sum of edge energies:

$$E_{\text{int}} = \sum_{n=1}^{N} E_{edge}(i,j), \tag{2}$$

in which i is the node index of the first node and j is the node index of the second node of the edge n of the total N edges. The *external energy* is defined as the sum of the values of the local maxima in Ψ that are associated with the hypothesis graph:

$$E_{ext} = \sum_{m=1}^{M} \Psi(\bar{p}_m), \tag{3}$$

in which $\Psi(\bar{p}_m)$ is the value for Ψ at the position $\bar{p}_m$. The total hypothesis energy is defined as:

$$E_{tot} = E_{\text{int}} - E_{ext} \tag{4}$$

The allowed deviation of hypothesis graph from the model graph can be controlled by changing the edge value K. By increasing K_i the penalty for deforming the edge i is increased. The described similarity measure provides flexibility by considering object probabilities and their geometry at the same time. The matching algorithm aims to find the hypothesis that optimally matches the model graph and the total hypothesis energy is calculated for all of the hypotheses graphs. The hypothesis with the least total energy is selected as the optimum.

2.4 Discussion of Method Properties

The edge energy itself is independent of its orientation even if the matching edges represent a different orientation, because it only depends on the absolute difference in edge length (See Figure 3). However, a hypothesis graph that differs only from the model graph by the orientation of one or two edges does not become a good match to the model graph per se. The allowed deviation in the orientation of an edge in the model graph depends on the amount of edges that connect to either one of its nodes. For example, a node that has only a single edge, can be positioned anywhere on a circular path from the opposing node without incurring any energy penalty. A node that is connected to two other nodes does not have this property. Wrongly oriented edges that are connected to nodes that have at least two edges will always cause edge length differences in the hypothesis graph, so that its internal energy is increased. The energy based graph matching algorithm is fundamentally rotational invariant, but the hypothesis generation step causes only hypotheses that are just slightly rotated versions of the model graph to be evaluated at all. Rotational invariance could be achieved within the matching framework by altering the hypothesis generation step or by searching for rotated versions of the model graph.

3 Results

The described method was implemented and tested. In this section results are presented. First a method for evaluation is proposed. Second, a detailed example of applying the algorithm is shown. Finally, the results of testing the effects of various parameters are described.

3.1 Evaluation Method

The detection performance is tested under influence of varying amount of clutter using the following evaluation method. A near-regular set of local maxima using r rows and c columns is created synthetically by the addition of noise from a Gaussian distribution $G(\mu_p, \sigma_p)$ to the nodes in a $r \times c$ grid. Each node has a probability value which is randomly drawn from another Gaussian distribution $G(\mu_t, \sigma_t)$. The resulting positions and probabilities correspond to the true object attributes. Correctly detected objects are termed true positives, whereas missed objects (false negatives) contribute to the false reject rate (FRR). Ideally the FRR is equal to zero, meaning that all true object positions were found. The ability of the algorithm to hypothesize locations for missing objects is tested by removing local evidence maxima before applying the method for detection. Furthermore, to test the sensitivity of the method to noise two aspects have to be tested. First, false positives are added, where their number is chosen in relation to the number of true positives. Secondly, the positions and probabilities of the false positive detections are randomly drawn from separate Gaussian distributions. The parameters of these Gaussian distributions can be varied to test the influence of noise in the detector output on the performance of the method.

3.2 Evaluation Results on Synthetic Data

The synthetic dataset comprised images containing a 3 by 5 grid of object probability maxima. The inlier acceptance distance was set to 10 pixels. The mean FRR was determined over five hundred evaluation runs. In Figure 4 the process of matching and dataset generation is illustrated. Figure 4 left shows the model graph. In the next image a set of peaks is shown, representing evidence maxima with a nearly regular structure. The next image shows one of the evidence maxima removed to simulate missing data. The next image shows the effect of added noise. On the far right the resulting optimal match from applying the algorithm is displayed. In this case the correct local maxima are found. The location of the missing object is estimated well.

The influence of the edge energy weight K. In this test the false positive rate was selected equal to the number of true positives. The probability value for true positives were set to 1. The probability value for the false negatives were drawn from a Gaussian distribution with a mean of 1 and a standard deviation of 0.5. The positions of the true positives were picked from a Gaussian distribution with a mean value of the correct position of the true positive and a standard deviation of 10 pixels, meaning that 95% percent of the true positives were displaced by up to 20 pixels from the correct true positive positions. This corresponds to a deviation of 13% and 4% with respect to the node distances in the row and column directions of the grid respectively. The edge parameter

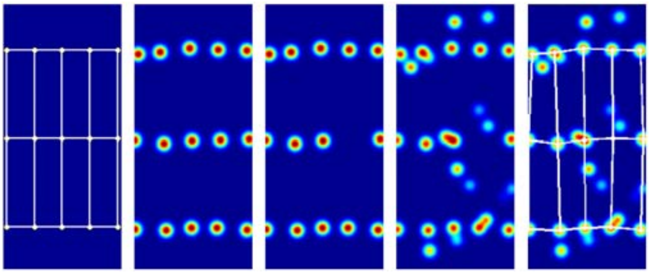


Fig. 4. Probability distribution of true positive object detections (left), with added noise (middle) and the optimal hypothesis graph found (right)

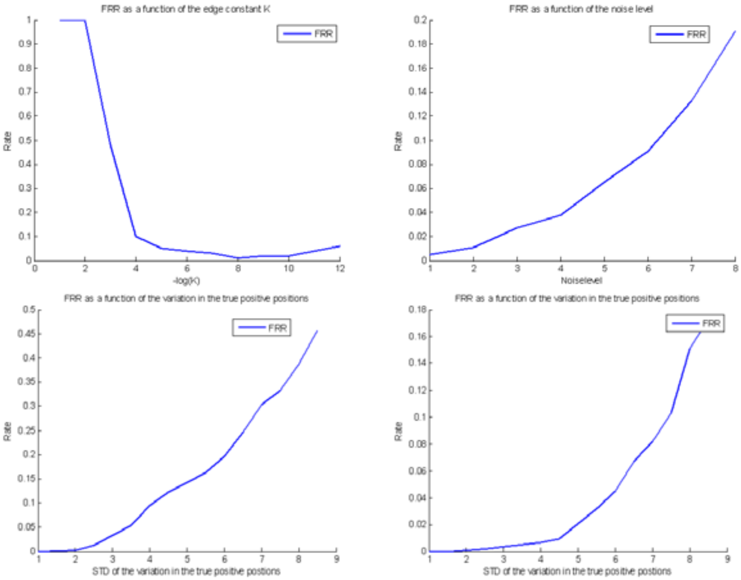


Fig. 5. The FRR as a function of the spring constant K (top left). The FRR as a function of the noise level (top right) and the FRR as a function of the variations of the true positive positions for a noise level of 5 (bottom left) and a noise level of 1 (bottom right).

K was varied between 100 and 10^{-10}. In Figure 5 (top left) the average FRR over five hundred runs is plotted. It can be seen that the FRR is reduced to a minimum at a value of 10^{-8}. For lower values of K the model graph can be deformed too freely, matching the nodes to local maxima that are not true positives. For higher values of K the model graph can not be deformed enough so that it fails to match to the correct local evidence maxima.

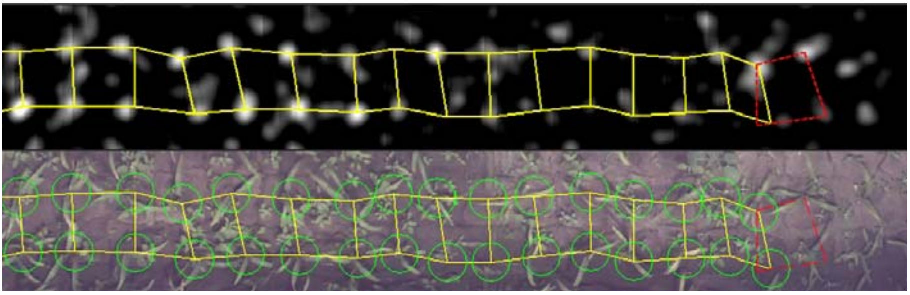


Fig. 6. Some examples of near regular object configurations

The influence of the noise level. The noise level is defined as the ratio between the number of true positives and the number of introduced false positives. The edge constant K was set to 10^{-7}. The other parameters were equal to the previous test. In Figure 5 (top right) the average FRR over five hundred runs is plotted. The FRR is largely unaffected by the noise level until a noise level of about 5 for which the FRR is at 6%. As the amount of noise further increases to almost 20% for a noise level of 8.

The influence of variations in true positive positions. The near-regularity of object configurations is expressed in slight variations of the positions of the true positive detections. To test the influence of these variations on the ability of the algorithm to find the correct true positives, their positions are varied between one and 8.5 times the standard deviations. The noise level was set to 5. The other parameters were not changed from the previous experiment. In Figure 5 (top left) the average FRR over five hundred runs is plotted. At a standard deviation = 5 the FRR crosses the 10% mark. High noise levels increase the chance of node mismatches, so it is expected that the noise level heavily influences the tolerance to model variations. Reducing the noise level to 1, the FRR is reduced significantly.

Test on real images. A radial symmetry detector [9] was used to detect hypothetical centers of crop planted at a near regular structure. The method successfully detects the nearly regularly positioned vegetation to within a few pixels accuracy under challenging conditions of clutter, as an example shows in Figure 6.

4 Conclusion

In this paper a novel non-rigid matching technique is presented to reliably detect nearly-regular object configurations in images in presence of substantial clutter. The proposed approach deals with object absence and noise by using a flexible structural prior. The optimal spatial arrangement follows from computing the globally optimal solution to a non-rigid matching problem by minimizing an energy cost function. Evaluation shows that the method is robust to outliers (clutter) amounts equal to the amount of true positives. The robustness decreases as the noise level increases. The method successfully detects near regularly positioned objects even if some of these objects are missing.

References

1. Fischler, M., Elschlager, R.: The representation and matching of pictorial structures. IEEE TC 22(1), 67–92 (1973)
2. Widrow, B.: The rubber mask technique. Pattern Recognition 5(3), 174–211 (1973)
3. Amit, Y., Kong, A.: Graphical templates for model registration. IEEE PAMI 18(3), 225–236 (1996)
4. Felzenszwalb, P.F.: Representation and detection of deformable shapes. IEEE PAMI 27(2), 208–220 (2005)
5. DiMaio, F., Shavlik, J.: Belief Propagation in Large, Highly Connected Graphs for 3D Part-Based Object Recognition. In: Perner, P. (ed.) ICDM 2006. LNCS (LNAI), vol. 4065, pp. 845–850. Springer, Heidelberg (2006)
6. Cootes, T.: An Introduction to Active Shape Models. In: Baldock, R., Graham, J. (eds.) Image Processing and Analysis, pp. 223–248. Oxford University Press, Oxford (2000)
7. Wiskott, L., Fellous, J.M., Kruger, N., von der Malsburg, C.: Face Recognition by Elastic Bunch Graph Matching. In: Jain, L.C., et al. (eds.) Intelligent Biometric Techniques in Fingerprint and Face Recognition, ch. 11, pp. 355–396. CRC Press, Boca Raton (1999) ISBN 0-8493-2055-0
8. Kushal, A., Schmid, C., Ponce, J.: Flexible Object Models for Category-Level 3D Object Recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007, pp. 1–8 (2007)
9. Loy, G., Zelinsky, E.: A Fast Radial Symmetry Transform for Detecting Points of Interest. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 358–368. Springer, Heidelberg (2002)

Traffic Scene Segmentation and Robust Filtering for Road Signs Recognition

Bogusław Cyganek

AGH University of Science and Technology
Al. Mickiewicza 30, 30-059 Kraków, Poland
cyganek@agh.edu.pl

Abstract. The paper describes a method for automatic scene segmentation and nonlinear shape-preserving filtering for precise detection of road signs in real traffic scenes. Segmentation is done in the RGB color space with a version of the fuzzy k-means method. The obtained posterior probabilities are then nonlinearly filtered with the novel version of the shape-preserving anisotropic diffusion. In effect more precise detection of object boundaries is possible. Thanks to this, the overall quality of the detection stage was increased, as it was confirmed by many experiments.

1 Introduction

Automatic recognition of the road signs (RSs) finds still growing interest of researchers and the automotive industry. A reliable system of this type, if assembled into a modern car, could improve safety of the traffic, assisting a driver and informing about seen speed limits or expected obstacles on the road. However, such systems still need to improve their performance for reliable operation in the hard road conditions.

An interesting system for RS recognition was proposed by Escalera *et al.* [7]. It is divided into the detection and recognition modules. An input image is color segmented in the HSI color space. Then recognition is formulated as an energy minimization problem. Paclík *et al.* [11] propose a similar system with classifiers using a trainable similarity measure. Their system relies heavily on template matching, however these takes places in local areas selected by especially trained classifiers. The work provides also an interesting overview of many different types of classifiers in context of real RSs recognition. Another type of a system for signs recognition was proposed by Chen *et al.* [2]. A recent system by Bascón *et al.* [1] consists of three stages: segmentation, detection, and recognition. The segmentation is proposed to take place in the HSI space with a fixed threshold values. Detection stage starts with shape discrimination followed by the FFT to deal with change in rotations and scale. Finally, recognition is done with a set of multi-class SVM classifiers with the RBF kernels, operating on the intensity signals of pictograms.

However, in all RS recognition systems the crucial part is a reliable detection of the signs. This, in turn, is based on the precise segmentation of the images gathered in real traffic conditions. In this paper we address this problem.

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 292–299, 2010.

The paper presents a method of color segmentation and shape-driven post-processing module for reliable detection of the RSs. The first task is achieved with the fuzzy k-means method, whereas the second with a novel version of the anisotropic diffusion operating on the probability fields, driven however by the intensity signal of the input image. Its purpose is to substitute a simple HSI segmentation used in the already developed RS recognition systems [3]–[5].

The paper starts with a discussion of the architecture of the system. Then details of the fuzzy k-means method, set to the RS detection task, are presented. In section (4) the novel version of the anisotropic diffusion process, employed to the filtering of the probability fields obtained from the fuzzy k-means, is discussed. Finally, the experimental results and conclusions are provided.

2 Architecture of the System

Architecture of the proposed system for preprocessing of the traffic images is depicted in Fig. 1. The proposed module was designed to improve performance of the already developed modules for the road signs detection [3]–[5].

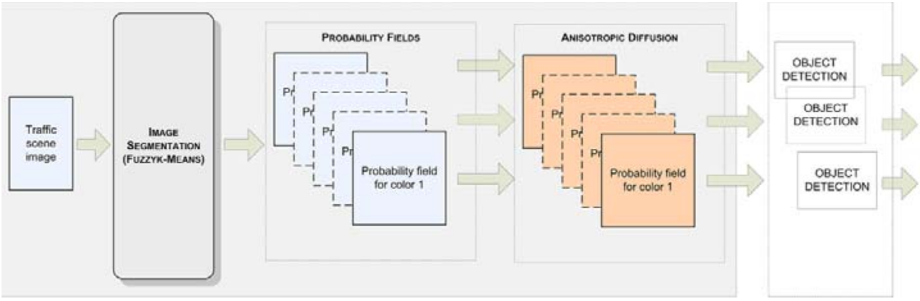


Fig. 1. Structure of the ensemble of classifiers for image segmentation

The preprocessing starts with the segmentation module which accepts an input color image and outputs a selected number of probability fields, each corresponding to the segmentation in respect to a single color value. In our realization the fuzzy k-means algorithm was used, although other methods can be also employed. As it will be discussed further on, the probability fields can be joined with other segmentation methods, producing their output in a form of a probability field. These can be simply multiplied, or a fuzzy *min* approach can be undertaken.

However, the obtained probability fields exhibit many deficiencies, especially for not well lightened or shadowed images [3]–[5]. Frequently the borders of the signs tend to be jagged which negatively influences their detection. We propose to improve quality of the probability fields with the novel formulation of the anisotropic diffusion filtering. This nonlinear filtering was first proposed by Perona *et al.* [12] and allows image smoothing which preserves object edges. In our formulation the anisotropic diffusion is applied to the probability field, as originally proposed by Sapiro [14]. However, in our system a novel control function, directly driven by the magnitude of the image intensity signal, instead of a gradient of probability, is proposed. Such an approach results in improvements in quality of detection. Finally, in a discrete realization of the partial differential equation for the anisotropic diffusion we propose to use the interpolating filters for efficient and stable differentiation of discrete signals.

3 Input Space Segmentation

Our main idea is to simultaneously segment the input image into a number of clusters determined by the number of colors expected in that image and which are characteristic to the expected road signs. These are: red, green, blue, and yellow, as well as black and white. For segmentation we use the fuzzy k-means algorithm due to its robustness and good convergence properties [8][9].

Given an input image the algorithm starts with choice of the initial number of clusters, for each a mean value μ_i is chosen. As already mentioned, these are the characteristic colors expected for the real road signs. In our experiments these were either 'pure' colors, such as (255,0,0) for red, etc., or mean values of the point samples manually gathered from the test images, as described in [5]. After initialization the method proceeds iteratively by assigning each point $\mathbf{x}_i$ to the closest mean μ_m in M clusters, as to minimize the following function

$$Q = \sum_{m=1}^{M} \sum_{i=1}^{N} P\left(c_m|\mathbf{x}_i\right)^{\gamma} \left\|\mathbf{x}_i - \mu_m\right\|_L, \tag{1}$$

where $||.||_L$ denotes a chosen distance metric

$$D_{im} = \left\|\mathbf{x}_i - \mu_m\right\|^2. \tag{2}$$

In the above M denotes a number of clusters, N is a number of data points. For a given set of means the probability values are computed in accordance with the following

$$P\left(c_m|\mathbf{x}_i\right) = \frac{\left(D_{mi}\right)^{1/(1-\gamma)}}{\sum\limits_{k=1}^{M} \left(D_{ki}\right)^{1/(1-\gamma)}}, \tag{3}$$

where γ is a parameter which should be greater than 1 (and in practice less than 5). When γ is close to 1, then the described algorithm approaches the classical k-means. If D_{ki}=0 for some indices k, i then $P(c_k|\mathbf{x}_i)$=1 and all other are 0. Finally, the new mean values are re-computed as follows

$$\mu_m = \frac{\sum_{i=1}^{N} P\left(c_m|\mathbf{x}_i\right)^{\gamma} \mathbf{x}_i}{\sum_{i=1}^{N} P\left(c_m|\mathbf{x}_i\right)^{\gamma}}. \tag{4}$$

The algorithm stops when there is no significant change in (4). In practice the convergence is reached in few dozens of iterations. After reaching the stable state, probability values (3) are assigned to each pixel of the initial image.

The simplest way to choose the initial values of the means in (2) is to set M=6 'pure' colors, which are characteristic of the road signs:
red (255,0,0), green (0,255,0), blue (0,0,255), yellow (255,255,0), black (0,0,0), and white (255,255,255).
Certainly M >6 can be also set if other objects with known characteristic colors are to be detected. However, this extends computations and memory occupation.

The probability values (3) for each class are then used to analyze the scene in the context of road signs recognition. Additionally the probabilities can be joined with responses from other detection modules, such as the fuzzy detector [3] or a Gaussian model built from the experimentally gathered characteristic points of the road signs [5]. Thanks to this an ensemble of classifiers with the multiplication rule is created. This usually shows more precise response than a single classifier [10].

4 Anisotropic Filtering of the Probability Fields

Anisotropic diffusion was proposed by Perona *et al.* [12] as a modification to image smoothing with the heat equation which preserves sharp object boundaries. The main idea is to use a control function $g()$ in the computation of the Laplacian that stops smoothing if an edge is encountered. This procedure can be written in the PDE form

$$\frac{\partial P\left(\mathbf{x},t\right)}{\partial t} = div\left[g\left(\|\nabla P\|\right)\nabla P\right], \tag{5}$$

where P is a signal to be smoothed, e.g. this can be the intensity signal if $P = I$.

The method found broad applications in image filtering, since it allows noise reduction without smearing boundaries of the objects. Then Sapiro proposed to use the anisotropic filtering to smooth out the posterior probabilities used in a process of object classification. As reported in [14] this increased precision of the classification based on the posterior probabilities. We propose a modification to this method, which can be used in the post-processing of the probability fields of the road scene segments found with the previously described methods. The modification relies on reformulation of the control function in such a way that it is influenced directly by the intensity signal of an image instead of the smoothed probability values. We also do not enforce probability normalization in each step since the probability values are positive and this values add up in each iteration of the filtering. Only after the process is finished, the resulting probability field is normalized once. The above equation takes the form

$$\frac{\partial P\left(\mathbf{x},t\right)}{\partial t} = div\left[g\left(\|\nabla I\|\right)\nabla P\right]. \tag{6}$$

This slight modification has significant consequences during operation. It allows spread of the probability values up to the real object boundaries instead of the boundaries of the probability region. Thanks to this the detected object with high probability values have also sharp boundaries which allow more reliable detection and classification by further modules of the system.

For the control function g Perona *et al.* [12] propose the $g(x)=(1+x^2/k^2)^{-1}$, where k is a positive constant. However in our realization the Tukey biweight function is used due to its superiority in leaving untouched strong signal variations, as shown by Sapiro [14]. It has the following form

$$g(x) = \begin{cases} \frac{1}{2}\left(1 - x^2/k^2\right)^2, & |x| \le k \\ 0, & otherwise \end{cases} . \tag{7}$$

For the parameter k in (7) the so called robust scale

$$k_r = 1.4826 \cdot med\left(\|\nabla I - med\left(\|\nabla I\|\right)\|\right) \tag{8}$$

was used which is computed from the monochrome version of the original image (and not from the probability field). The parameters used in (8) are explained in [14].

5 Experimental Results

The presented method was implemented in C++. Experiments with many real images were run on the computer with 2GB RAM and Pentium Core 2 T7600 @ 2.33GHz. Fig. 2 depicts two selected color test images of resolution 640x480 for which further results of our method are also presented.

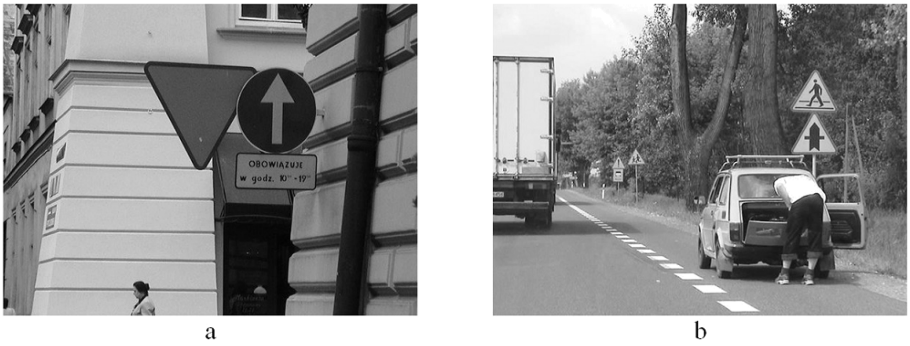


Fig. 2. Two real traffic scenes used in presented experiments

As alluded to previously, the initial values for the fuzzy k-means centers are the six colors characteristic of the road signs. The upper row of Fig. 3 shows three probability fields computed from the image in Fig. 2a in accordance with (3) for yellow, white and blue colors, respectively. It is evident that apart from the road signs other objects with similar colors have also high response. Our previously developed object detectors are to operate in these probability fields, searching for characteristic figures based on their salient points. However, due to noise and some some salient points cannot be determined reliably. Nevertheless, the probability maps can be improved with the proposed version of the anisotropic diffusion, which diffuses the probability values respecting object boundaries at the same time. The nonlinear diffusion process is controlled by the stopping function (7) driven by the gradient of the intensity signal. The robust scale (8) is also employed. In experiments its value reached from 5.5 up to 12.3. The number of iterations was chosen to be 125 to achieve good smoothing results. These are depicted in the lower row of Fig. 3. Under closer examination it can be observed that interiors of the objects get well smoothed while their boundaries are crisp. Thanks to this feature the detectors can more reliably detect the salient points of the signs.

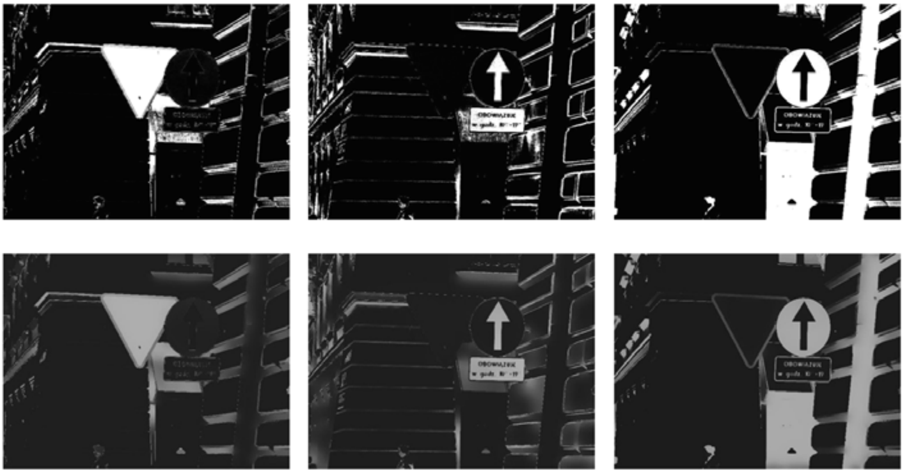


Fig. 3. Results of image segmentation with the fuzzy k-means for $\gamma = 1.4$: The probability fields (upper row). These filtered with our version of the anisotropic diffusion.

Fig. 4a shows results of yellow color segmentation of an image depicted in Fig. 2b with the method developed in [4]. It is visible that object boundaries are not sharp which impediments the triangle detection process. In contrast, the proposed method produces more precise segmentation which can be viewed in Fig. 4b and Fig. 4c. The latter version was obtained after 100 iterations of the anisotropic diffusion process (6) with conditions (7) and (8).

The proposed system of image segmentation and postprocessing not only improves precision but also the recall factor of the RS detection process. In our system the improvement was in order of 2-15%. The main drawback is execution time which, in our non-optimized setup, was in order of 20-32 s per image of resolution 640x480. This is due to iterative procedures both for the fuzzy

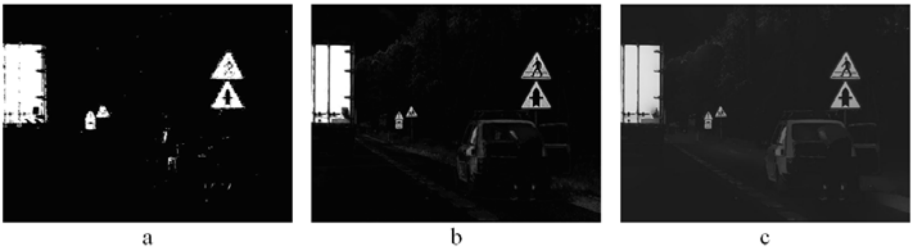


Fig. 4. Segmentation of the image in Fig. 2b for yellow objects. Results of the method from [4] (a). The probability field obtained with the proposed method with $\gamma = 2.2$ (b). The anisotropic filtered probability field (c).

k-means, as well as for the anisotropic diffusion. However, an improvement of performance is also possible since the latter can be performed in parallel for each of the probability spaces.

We also plan porting of some computations to the GPU card which can shorten computation time at least an order of magnitude.

6 Conclusions

In this paper we present a hybrid method of color image segmentation and shape-driven postprocessing for improvement in precision of detection of the road signs. Segmentation is performed with a customized version of the fuzzy k-means method, which starts with initialization of the characteristic six colors encountered in the road signs. In effect the six probability fields are obtained which for improvement are nonlinearly filtered with a novel version of the anisotropic diffusion. To account for the shape of real object the diffusion process is nonlinearly controlled by a robust function. It is directly driven by the module of the gradient of intensity signal, instead of the probability values which do not convey information on real object boundaries. Such novel approach allowed increase of the quality of RS recognition by the already developed recognition systems. This was verified by many experiments performed on real traffic scenes. However, the method was tested only in daylight conditions and it needs further improvements to speed up the computations.

Acknowlegement

This work was supported from the Polish funds for scientific research in 2010.

References

1. Bascón, S.M., Rodríguez, J.A., Arroyo, S.L., Caballero, A.F., L"opez-Ferreras, F.: An optimization on pictogram identification for the road-sign recognition task using SVMs. Computer Vision and Image Understanding 114(3), 373–383 (2010)
2. Chen, X., Yang, J., Zhang, J., Waibel, A.: Automatic Detection and Recognition of Signs from Natural Scenes. IEEE Trans. on Image Proc. 13(1), 87–99 (2004)

3. Cyganek, B.: Soft System for Road Sign Detection. In: Theory and Applications of Fuzzy Logic and Soft Computing. Advances in Soft Computing, vol. 41, pp. 316–326. Springer, Heidelberg (2007)
4. Cyganek, B.: Real-Time Detection of the Triangular and Rectangular Shape Road Signs. In: Blanc-Talon, J., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2007. LNCS, vol. 4678, pp. 744–755. Springer, Heidelberg (2007)
5. Cyganek, B.: Color Image Segmentation With Support Vector Machines: Applications To Road Signs Detection. International Journal of Neural Systems 18(4), 339–345 (2008)
6. Cyganek, B., Siebert, J.P.: An Introduction to 3D Computer Vision Techniques and Algorithms. Wiley, Chichester (2009)
7. de la Escalera, A., Armingol, J.A.: Visual Sign Information Extraction and Identification by Deformable Models. IEEE Transactions On Intelligent Transportation Systems 5(2), 57–68 (2004)
8. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. Wiley, Chichester (2001)
9. Gan, G., Ma, C., Wu, J.: Data Clustering. Theory, Algorithms, and Applications. SIAM, Philadelphia (2007)
10. Kuncheva, L.: Combining Pattern Classifiers. In: Methods and Algorithms, Wiley, Chichester (2004)
11. Paclík, P., Novovičová, J., Duin, R.P.W.: Building road sign classifiers using a trainable similarity measure. IEEE Transactions on Intelligent Transportation Systems 7(3), 309–321 (2006)
12. Perona, P., Malik, J.: Scale-Space and Edge Detection Using Anisotropic Diffusion. IEEE Trans. On Pattern Analysis and Machine Intelligence 12(7), 629–639 (1990)
13. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C. The Art of Scientific Computing, 3rd edn. Cambridge University Press, Cambridge (2007)
14. Sapiro, G.: Geometric Partial Differential Equations and Image Analysis, Cambridge (2001)

Fast Distance Vector Field Extraction for Facial Feature Detection

Wojciech Czarnecki, Szymon Gburek, and Maciej Smiatacz

Faculty of Electronics, Telecommunications and Informatics
Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland
{Wojciech.Czarnecki,Szymon.Gburek}@zak.eti.pg.gda.pl,
slowhand@eti.pg.gda.pl

Abstract. This work is related to the method of facial feature detection based on distance vector fields (DVFs), recently proposed by Asteriadis *et al.* We briefly present the concept and describe improvements that we introduced to the original solution. The main advantages of our approach are the reduced computational complexity of the DVF extraction algorithm as well as the enhanced precision of the resultant vector field.

1 Introduction

Face recognition has recently become one of the most important biometric technologies. After many years of research and spectacular failures algorithms finally turn into mature and stable solutions. Latest reports from automatic face recognition competitions indicate that in particular cases computer systems perform better than humans [1]. However, in terms of speed, accuracy and practical usefulness of this technology there is still much to improve.

Every biometric application based on face recognition must perform the localization of the image fragment that will be further analyzed. Consequently, during the last 15 years numerous face detection methods have been proposed. Most of them were highly unreliable and provided acceptable results only in controlled environment (uniform background, good illumination conditions, single upright face photographed with high quality camera). The situation changed radically after the publication of the paper by Viola and Jones [2]. The authors developed extremely fast and accurate localization algorithm based on Haar features, AdaBoost concept and hierarchical cascaded classifier. Their solution, performing very well in practical applications, became a standard for face localization. Although improvements are still available, the problem of automatic face detection can be considered solved.

The AdaBoost method almost perfectly extracts rectangular regions containing face images. Recognition algorithms, however, usually require proper localization of facial features (typically eyes and mouth areas) as well. Several approaches can be used to achieve this, depending on computational complexity restrictions and accuracy level needed. In our previous research we tested the Active Shape Models and obtained moderate results [3]. Currently we are

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 300–307, 2010.

working on face recognition system for mobile devices, so the low complexity of the algorithm is essential. Therefore we decided to examine the method based on distance vector fields, developed by Asteriadis *et al.* [4], which appeared to be relatively straightforward. In the following sections we briefly describe the original concept and our modifications that improved the performance of the facial feature localization process.

2 Distance Vector Fields for Facial Feature Localization

The concept of assigning two integer values to each pixel, thus forming a specific vector field of an image, appeared in the work of Danielsson [5], discussing efficient ways of calculating Euclidean distance mapping. The term distance vector field (or DVF), however, was rarely used before the paper on facial feature detection by Asteriadis *et al.* [4] appeared. The general idea is to find the closest object pixel (k, l) for each pixel (i, j) belonging to the background of the binary image. In other words, we have to calculate

$$\mathbf{v}(i,j) = [k-i, l-j]^{\mathrm{T}}, \quad (i,j) \in B, \quad (k,l) = \arg\min_{m,n \in O} D((i,j),(m,n)) \qquad (1)$$

where D is a distance metric and O, B represent sets of object and background pixels. Authors of [4] noticed that DVF can be used to describe the geometry of an object, a human face for example. In this case the edge map of an image must be extracted first, using Canny algorithm [6], so that the set O will contain only the edge pixels. In this version each element of the distance vector field, calculated for the given (i, j) coordinates, contains a vector pointing to the closest edge pixel. Thus the DVF encodes geometrical relations between image pixels and their neighboring edges. In [4] the use of the edge map instead of pixel intensity values is presented as an additional advantage offering a certain level of invariance to brightness variations and poor lighting conditions. In reality this is not so obvious, because adverse lighting can seriously affect the quality of edge map, causing some important feature contours to disappear. Moreover, the performance of the final algorithm relies on proper selection of the parameters of Canny method.

The general concept of DVF is illustrated on Fig. 1a. The contents of the DVF can be presented as the two grayscale images formed from the horizontal and vertical components of the vector field (see Fig 1bc).

Apart from the calculation of the DVF, the method of facial feature localization described in [4] contains the following steps: 1) eye region detection, 2) mouth region detection, 3) eye center localization, 4) mouth corner localization.

The above operations are mostly based on template matching supported by some heuristic rules describing the geometry of a human face. To obtain the eye region template the DVFs of 58 manually extracted right and left eye images scaled down to the dimensions of $X = 26$ by $Y = 26$ were used. Similarity between the scanned image fragment $\mathbf{V}$, described by its DVF vectors $\mathbf{v}_i$, and the template $\mathbf{M}$, represented by its DVF vectors $\mathbf{m}_i$, is inversely proportional to the distance defined as

$$E_1(\mathbf{V},\mathbf{M}) = \sum_{i=1}^{XY} |\mathbf{v}_i - \mathbf{m}_i| = \sum_{i=1}^{XY} \sqrt{|v_{ix} - m_{ix}|^2 + |v_{iy} - m_{iy}|^2} \tag{2}$$

where v_{ix}, v_{iy} are the horizontal and vertical components of vector $\mathbf{v}_i$. Analogous approach was followed to detect mouth region. This time 16 regions with dimensions $X = 36$ by $Y = 16$ were extracted to form the DVF template. Since the mouth localization is difficult because the skin color is often similar to the color of lips and additional distortions of the DVF caused by the presence of beard may appear, an extra factor was included in formula (2):

$$E_2(\mathbf{V},\mathbf{M}) = E_1(\mathbf{V},\mathbf{M}) + \bar{E}_1 \frac{\bar{I}_{HE}}{I_{HE}} \tag{3}$$

where I_{HE} denotes the number of pixels in the horizontal edge map of $\mathbf{V}$, $\bar{E}_1$ is the mean value of $E_1(\mathbf{V},\mathbf{M})$ calculated for all regions $\mathbf{V}$, i.e. for all 36×16 windows that fit within the candidate mouth area, and $\bar{I}_{HE}$ is the mean value of I_{HE} over all image fragments $\mathbf{V}$.

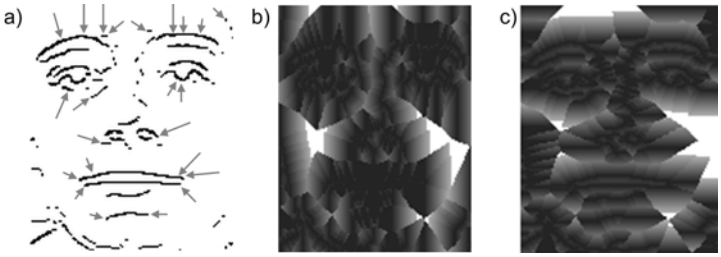


Fig. 1. a) Sample elements of a distance vector field, i.e. the vectors pointing to the nearest edge pixels, b) horizontal components of the DVF, c) vertical components of the DVF (darker regions correspond to smaller values).

The eye center localization is performed in four steps.

1. Reflections are removed. Because specular highlights often appear in eye images and may disturb the localization process, the small bright areas occupying less than 1% of pixels are filled with the average intensities of their neighborhoods. The algorithm is briefly presented in [4], the detailed description of our implementation is given in the next section.

2. Projections of vertical and horizontal components of image gradient are constructed. Four largest values in each projection are selected and the central peaks define an initial estimate of the eye center inside the rectangular eye region.

3. The darkest row and the darkest column is found within a small area around the point selected in step 2.

4. The darkest region of size $0.25X_{eye} \times 0.25Y_{eye}$ is searched for around the point found in the previous step. The middle of this area defines the eye center.

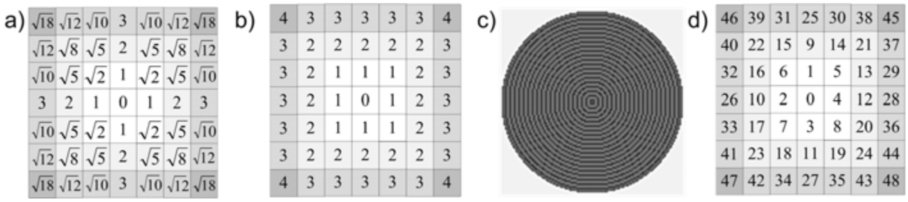
Mouth corners are localized using color information. Although our previous experiments related to face localization in color images [7] led us to the conclusion that in practice stable results cannot be obtained this way, in case of the small preselected mouth region segmentation based on hue values can be reliable. First the Otsu algorithm [8] for automatic thresholding is employed and then localization is performed by looking for the rightmost and leftmost point in the class of pixels which has the mean hue value closer to 0°. All the foreground components whose area is smaller than 15% of the mouth region are removed.

3 Improvements of the DVF Method

The key element of the facial feature extraction approach described in [4] is the algorithm responsible for calculation of the distance vector field – this is why we will call it a DVF method. Particular care must be taken when implementing the DVF creation process as it is potentially the most time-consuming and computationally complex fragment of the whole application code. The authors of the original publication, however, addressed this problem in a way that may raise doubts regarding the efficacy of their solution. They notice that previously, for example in [5], DVFs were mainly used only as an *intermediate* step for calculation of the distance map. Thus, it is evident that it is not necessary to create the distance map before the DVF extraction. On the other hand authors declare that in their implementation DVFs are *derived from* the distance map of the edge image. Consequently, the distance map is prepared first using the method described in [9] and so their approach is certainly not optimal as the redundant procedure is carried out at the beginning of the process.

In the original algorithm the 8-neigborhood of each pixel (i, j) is examined and the neighbor with the smallest distance map value is selected as the current pixel. Then the search is performed in the neighborhood of the current pixel and the whole procedure continues until the pixel with value 0 on the distance map is reached.

We found the above approach ineffective and decided to implement another concept – instead of starting the search from each image pixel we iterate over edge points and look for the background pixels for which the given edge point is the closest one. First we create a list of points defining the search order around each edge pixel. This is a list of rings with ascending radii determined by calculating



a)

$\sqrt{18}$	$\sqrt{12}$	$\sqrt{10}$	3	$\sqrt{10}$	$\sqrt{12}$	$\sqrt{18}$
$\sqrt{12}$	$\sqrt{8}$	$\sqrt{5}$	2	$\sqrt{5}$	$\sqrt{8}$	$\sqrt{12}$
$\sqrt{10}$	$\sqrt{5}$	$\sqrt{2}$	1	$\sqrt{2}$	$\sqrt{5}$	$\sqrt{10}$
3	2	1	0	1	2	3
$\sqrt{10}$	$\sqrt{5}$	$\sqrt{2}$	1	$\sqrt{2}$	$\sqrt{5}$	$\sqrt{10}$
$\sqrt{12}$	$\sqrt{8}$	$\sqrt{5}$	2	$\sqrt{5}$	$\sqrt{8}$	$\sqrt{12}$
$\sqrt{18}$	$\sqrt{12}$	$\sqrt{10}$	3	$\sqrt{10}$	$\sqrt{12}$	$\sqrt{18}$

b)

4	3	3	3	3	3	4
3	2	2	2	2	2	3
3	2	1	1	1	2	3
3	2	1	0	1	2	3
3	2	1	1	1	2	3
3	2	2	2	2	2	3
4	3	3	3	3	3	4

c)

d)

46	39	31	25	30	38	45
40	22	15	9	14	21	37
32	16	6	1	5	13	29
26	10	2	0	4	12	28
33	17	7	3	8	20	36
41	23	18	11	19	24	44
47	42	34	27	35	43	48

Fig. 2. a) A map containing Euclidean distances from pixel 0, b) subsequent rings around 0, c) visualization of 60 rings, d) the search order around 0.

the floor of the Euclidean distance from the center. Fig. 2abc shows the distance map, the subsequent rings around the 0 point and the visualization of the 60 rings used in one of our experiments. It can be seen that the higher-order rings approximate circles quite accurately. Each ring, however, may contain points with different Euclidean distances from central point 0, for example the 1^{st} ring contains points which are 1 or $\sqrt{2}$ units away from the center. Therefore, the points in each ring are sorted by their true distance values. For the points with the same distance value some arbitrary sequence can be applied. The final search order used in our application is presented on Fig. 2d.

The search rings can be created once and stored in the source code in a form of a table. Diverse approximations of discrete circles may be considered, resulting in slightly different distance vector fields. In this case, however, small imperfections in DVF calculation are negligible due to the nature of template matching algorithm. The speed of the code is the main concern, especially when the method is supposed to be applied to mobile devices. On the other hand, the procedure used in [4] is also prone to some errors.

In our approach during the first iteration all the edge points check if pixel no 1 in their search maps is a background pixel or another edge point. In case of a background pixel its label is changed from '-1' to the label e_i of the edge point performing the check. This way the closest edge point is assigned to the background pixel. In second iteration all the edge points check pixel no. 2 but now there is one more option: the pixel can belong to an edge, can be a background pixel or can be already labeled by another edge point during the previous iteration. Only in the second case the checked pixel can be assigned to the edge point e_i. In following steps the subsequent pixels indicated by the search map are processed. The loop ends when there are no pixels marked with '-1' in the image, i.e. when the nearest edge points were found for all background pixels.

The idea of ring was introduced to allow for fast elimination of the edge points that no longer participate in the creation of the distance vector field. In other words, the edge point that was not able to label any background pixel during the whole cycle defined by the current ring on the search map can be treated as inactive in the following iterations because there is no chance that the point will label any pixel in further cycles. The faster the inactive edge points are identified, the higher the speed of the algorithm.

To sum up, our algorithm assigns the closest background pixels to each edge point using predefined search order. After each cycle, when all the pixels belonging to the current ring around each edge point have been checked, the edge points that did not label any background pixel are marked as inactive and eliminated from further search. The calculation of distance map is not necessary.

Apart from improving the performance of DVF calculation algorithm we precisely defined the method of reflections elimination playing important role in the process of eye center localization. The procedure is as follows.

1. First the binary image of an eye is obtained with the help of the global thresholding algorithm by Otsu [8]. The reflection areas are represented by small white regions occupying less than 1% of the image.

2. In order to find the regions of interest the binary image is converted to 32-bit format. In one iteration all white connected components are filled with subsequent colors starting from 1 (0x00000001) and number of points in every connected component is determined. After this step there are no white areas in the 32-bit image which can be treated as a “region map”.
3. Using the information gathered in the previous step we are able to select small non-black connected components covering less than 1% of the eye region. These reflection areas must be filled with the average luminance of their neighborhoods in the original image. The color of the neighborhood for each of those small regions is determined by scanning the image across rows and columns and recording the luminance values of such pixels of the original image that correspond to the black pixels in the region map, which are adjacent to the reflection areas.
4. Each reflection area is filled with the average luminance of its neighborhood calculated in the previous step.

As we already mentioned, the search for mouth corners is carried out by looking for the first and the last pixel along x axis in the region extracted by hue thresholding. Only the hue values from the range of $[340°, 20°]$ are taken into account and the Otsu method [8] is used once again to find the class of pixels with hue values closer to 0°. The histogram, however, is not constructed directly from hue values but from the distances from each hue to 0°. The distance between hue values a and b is measured as follows:

$$d(a, b) = \pi - |\pi - |a - b|| \tag{4}$$

It means that, for example, the number of pixels with hue value of 355° is added to the number of pixels with hue value of 5° to form the 6^{th} bin of the histogram.

4 Experiments

We implemented the original DVF method by Asteriadis *et al.* [4] and our modifications described in Section 3 in a C# application that allowed us to experiment with different versions of algorithms calculating the distance vector fields. The code from *AForge* library was employed to perform typical tasks such as thresholding or edge detection. Because we are looking for the facial feature localization method suitable for mobile devices, the low computational complexity of the algorithm was our priority.

The problem of DVF extraction is closely related to the complex task of calculating the Euclidean distance transform of an image. The comparative study of appropriate algorithms can be found in [10]. We did not investigate all of them, but the solution that we proposed is more efficient than the method presented in [4].

First, our approach does not require the initial calculation of the distance map. Although, according to results reported in [10], for small images with the size of 100 × 100 the time of distance map computation is negligible on a 2GHz computer, it can be an important factor in the case of larger photographs and less efficient mobile devices.

Moreover, the speed of distance transform algorithms highly depends on the contents of the image (randomly distributed pixels or straight lines etc.). We face the same issue in the case of DVF computation: in the typical situation when the number of edge points is relatively small our method is considerably faster than the algorithm described in [4]; however, for dense edge maps the benefits of our approach become less evident. We compared the performance of the original DVF calculation algorithm and the fast version that we have proposed using the set of 7254 face images from FERET database [11]. The results are presented in Table 1. For 256 × 384 photographs our method appeared to be 3 times faster even though the time required to create the distance map was not included in the value measured for the original algorithm. Our method is also much more accurate: the percentage of pixels for which the DVF values are incorrect is small and the errors are never greater than 1 pixel.

Table 1. The comparison of average results obtained for DVF [4] and Fast DVF algorithms on 7254 face images from FERET database

	DVF	**Fast DVF**
Computation time (ms)	810	246
Percentage of points with incorrect DVF values	31.2%	0.18%
Percentage of points with errors greater then 1 pixel	6.77%	0%

Although performance was our main concern, obviously it is the precision of feature localization that makes the algorithm useful in practice. So far we cannot confirm the opinion of the authors of [4] that “the method gives very accurate results, failing only in extreme cases”. We have not, however, performed large-scale quantitative tests on representative databases yet. Some examples of feature detection provided by our implementation are depicted on Fig. 3.

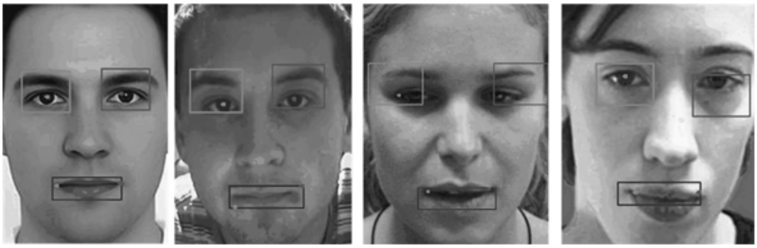


Fig. 3. Facial feature localization results obtained for selected face images

The intrinsic weakness of the method comes from the fact that the localization is entirely driven by heuristic rules and simple matching of fixed-size templates. Therefore, the accuracy of the algorithm extracting faces from the background (e.g. [2]) is crucial. If the face area cropped at this stage is slightly too small and does not include the whole head, the DVF algorithm fails because the eye and mouth templates become too small to match the image properly.

5 Conclusion

We have implemented and tested the newly proposed method of facial feature detection based on the distance vector fields [4]. We improved the algorithm by reducing its computational complexity and improving the accuracy of the DVF extraction procedure. Our experiments showed, however, that the results provided by the method are not stable nor accurate enough for real-life applications. Therefore in our further research we will focus on introducing some kind of elastic template matching to improve the localization accuracy.

Acknowledgments. This work has been supported by the Polish Ministry of Science and Higher Education as the research project no. N N516 36793 and by departmental grant no. 018906.

References

1. Phillips, P.J., Scruggs, W.T., O'Toole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L., Sharpe, M.: FRVT 2006 and ICE 2006 Large-Scale Results. NISTIR 7408 National Institute of Standards and Technology, Gaithersburg (2007)
2. Viola, P., Jones, M.J.: Robust Real-Time Face Detection. Int. J. Comp. Vision 57(2), 137–154 (2004)
3. Smiatacz, M., Malina, W.: Active Shape Models in Practice. In: Kurzyński, M., Puchala, E., Woźniak, M., Żolnierek, A. (eds.) Computer Recognition Systems. Advances in Soft Computing, vol. 30, pp. 451–459. Springer, Heidelberg (2005)
4. Asteriadis, S., Nikolaidis, N., Pitas, I.: Facial Feature Detection Using Distance Vector Fields. Patt. Rec. 42, 1388–1398 (2009)
5. Danielsson, P.E.: Euclidean Distance Mapping. Computer Graphics and Image Processing 14(3), 227–248 (1980)
6. Canny, J.: A Computational Approach to Edge Detection. IEEE Trans. PAMI 8(6), 679–698 (1986)
7. Smiatacz, M.: Practical Evaluation of the Basic Concepts for Face Localization. In: Computer Recognition Systems 2. Advances in Soft Computing, vol. 45, pp. 52–59. Springer, Heidelberg (2007)
8. Otsu, N.: A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. Systems, Man, and Cybernetics 9(1), 62–66 (1979)
9. Breu, H., Gil, J., Kirkpatrick, D., Werman, M.: Linear Time Euclidean Distance Transform Algorithms. IEEE Trans. PAMI 17(5), 529–533 (1995)
10. Fabbri, R., Da, F., Costa, L., Torelli, J.C., Bruno, O.M.: 2D Euclidean Distance Transform Algorithms: A Comparative Survey. ACM Comput. Surv. 40(1), 1–44 (2008)
11. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET Evaluation Methodology for Face Recognition Algorithms. IEEE Trans. PAMI 22(10), 1090–1104 (2000)

Vision-Based Vehicle Speed Measurement Method

Witold Czajewski and Marcin Iwanowski

Warsaw University of Technology
Institute of Control and Industrial Electronics
ul. Koszykowa 75, 00-662 Warszawa, Poland
W.Czajewski@isep.pw.edu.pl, iwanowski@ee.pw.edu.pl

Abstract. This paper introduces a novel method for vehicle speed detection based exclusively on visual information. The proposed system consisting only of a digital camera and a computer is able to identify both the speed of the passing vehicle and its licence plate numbers, which makes it an interesting alternative to existing and expensive photoradar systems. The principle of operation is simple: vehicles are identified by their licence plates and their speed is measured based on the vertical difference of their position in consecutive images. An experimental evaluation shows the high accuracy of vehicle speed measurement, comparable to the one provided by comercially available radar-based systems.

1 Introduction

In the contemporary world we observe a growing need for intelligent traffic management systems that would cope with the constantly increasing traffic. According to the statistics [8], in Poland in 2005 approximately 75% of all people and cargo transportation was the road transportation. Unfortunately the number of vehicles was growing much faster then the infrastructure, which led to the road system overload, higher environment pollution, larger accident number etc. According to various studies, an intelligent traffic system can increase the throughput by 20%, reducing at the same time pollution and acident number.

One of the key issues of traffic monitoring is the automatic vehicle speed measurement. Currently it is done by means of photoradars, which use radar or laser technology for speed measurement and may use image processing techniques for licence plate identification. Although these devices usually perform well, they are expensive and thus cannot be widely used, especially in less economically succesfull regions, where the disproportion between the increasing traffic and the infrastructure is particularly large.

The main contribution of this paper is a novel approach to speed detection that does not require any sophisticated and expensive hardware and uses only visual information both for licence plate recognition and speed measurement. The underlying idea of the proposed method is to use the geometrical information of the camera-road system in order to estimate the velocity of passing vehicles.

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 308–315, 2010.

The paper consists of 5 sections. In Section 2 the previous works are reported. In Section 3, the proposed approach is described. Section 4 contains experimental results, and finally, Section 5 concludes the paper.

2 Previous Works

Vehicle speed measurement methods based on video sequences were proposed e.g. in [7,16,4], an interesting method for estimating speed from a single, blurred image, in [5]. More general studies on motion in digital images can be found in [19,18]. The motion and object tracking in traffic scenes was the subject of numerous publications, just to mention [17,11]. In the literature on car license plate recognition (LPR) one can find various approaches. In [14,9] a method based on edge detection on graylevel images combined with adaptive thresholding and template matching is presented. The method described in [13] processed edges on color input image using fuzzy-logic approach to localize license plate while topological features and neural network to recognize characters. Paper [6] describes a method that applies Maximal Stable Extremal Regions to localize the plate and SVM-classifer the the second step of processing. In [2] the proposed algorithm made use of feature-salience theory and applied Hough transform to localize plate and probabilistic classifier in the OCR phase. Statistical approach was also applied in [1]. The morphological approach to license plate localisation can be found in [15,12,3].

3 Proposed Method

The underlying idea of the proposed method is to use the geometrical information of the camera-road system in order to extract the distance covered by the observed vehicles between consecutive videoframes and hence calculate the passing vehicles velocity. This concept, followed by considerations about the image acquisition is described in Section 3.1. Next, the licence plate detection and recognition algorithm is used. For better and faster results, image masking, contour analysis and grammatical analysis is applied as shown in Sections 3.2 and 3.3. Finally, Section 3.4 describes the speed measurement method in detail.

3.1 Camera Setup and Image Acquisition

Usually photoradars are mounted on the side of the road and cover a relatively small sector so as to photograph vehicles in their direct vicinity. The camera is triggered by the radar detecting an incoming vehicle and a photo is taken from a short distance. In our case, however, such a configuration is not possible as we have no other triggering device but a single camera. The only way of detecting moving vehicles is continuous image registration and analysis. In order to measure the velocity we need at least two images of an approaching car. If am inclined camera is mounted centrally over the road like shown in Fig. 1(a) it will register images like the one shown in Fig. 1(b).

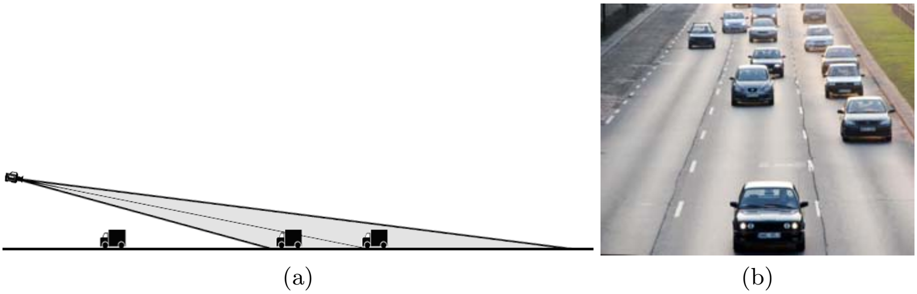


Fig. 1. The camera mounted over the road and looking at the incoming traffic (a), the view from the camera (b)

Now, in order to establish the distance to the observed vehicles one could utilize the fact that the licence plate is a rectangular object of a well known size. Finding its localization and orientation in space [10] would require however a very precise pinpointing of the plate's corners, which is not possible at far distances due to small resolution of the image. Another method could be based on the licence plate area measurement, but the results would be only slightly better for the same reason.

In our method we use a much more precise mechanism (see Section 3.4 for details). It is based on the observation that with the inclined camera, vehicles that are far away appear at the top of the image and they move down in the image as they approach. Establishing this nonlinear relation between the image location and camera-vehicle distance allows for speed measurement with sufficient precision.

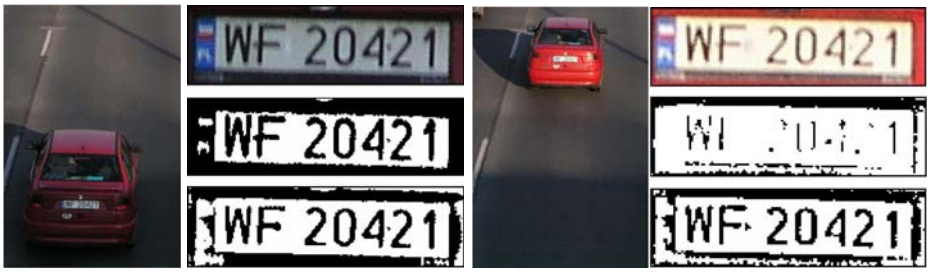


Fig. 2. Example of licence plate thresholding. The images show a car captured in shadow and sunshine. By side presented are the extracted plates (top) and effects of constant thresholding (middle - notice the right plate is unreadable) and adaptive thresholding (bottom - both plates are readable).

Since the proposed system is based exclusively on image analysis the resolution and overall quality of the acquired images is crucial. In our tests we used 5 and 6 megapixel images taken with a bright lens camera. This allowed for shorter exposure time while maintaining good image brightness, contrast and sharpness at low ISO rates. Images were taken at approximately 2 fps, which was sufficient for capturing at least 2 images of the same car even at high speeds and left 500 ms for image processing. That would be enough for a fast PC, but probably not for an embedded solution. The pictures were taken during the day in different weather conditions (from sunny to complete overcast).

3.2 Licence Plate Detection

Majority of LPR methods methods work best in situations when the expected size of a licence plate in an image is known a priori. This is the case of static or roadside photoradar LPR systems where the camera-car distance is almost constant and so is the size of the number plate. In our case, however, the linear dimensions of the licence plate doubles or even triples as cars are approaching the camera, which makes the plate detection algorithms more complicated.

Initially in our experiments we used an edge detection and vertical projection method on images with enhanced contrast. Without the enhancement many dirty and thus low-contrast plates were not found. Unfortunately, apart from succesfully detecting the licence plates, this method also returned many false positives such as car body parts with vertical details, gratings, fences, other letterings etc. This was caused by the character of the acquired images that contained many objects of different scale. In the end, all the found regions had to be filtered in order to exclude the non-plate regions. This was done by character analysis as described in section 3.3.

It turned out later during the experiments, that the initial detection step is completely useless. A horizontal series of neighboring characters defines a licence plate not worse than vertical edges used beforehand, but at the same time it returns the extracted characters that can immediately undergo recognition. The overall processing time was also shorter, which was the most important advantage of this approach.

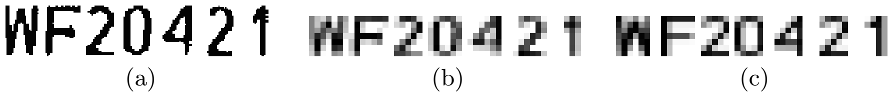
Finally, the plate detection algorithm based on character detection consisted of the following steps:

1. Adaptive thresholding of the entire image - we used different thresholding parameters for different sections of acquired images experimentally tuned to the expected size of the observed number plates. The adaptive thresholding algorithm was particularly succesful in correct segmentation of licence plates in different lighting conditions as opposed to a constant thresholding method (see Fig. 2).
2. Contour extraction from the binary image obtained in step 1 and analysis - rejection of all the contours that do not fall within experimentally defined bounds for size and height to width ratio.
3. Contour groupping - all the contours from step 2 are groupped with their horizontal neighbors. If a contour has too few or too many neighbors of similar size the entire group is rejected.

As the result we obtain not only areas containing licence plates but also extracted characters ordered from left to right belonging to separate plates.

3.3 Character Recognition

The next step of processing is the character classification. All the characters from a single plate are converted into grayscale and scaled down to 7x7 matrix each and normalized as shown in Fig. 3. This particular size was chosen after a number of experiments as it provided the best recognition results. Smaller number of divisions yields too generalized representation and bigger number results in too detailed representation of a particular character.



(a) (b) (c)

Fig. 3. Extracted characters in original resolution (a) and downsampled to 7x7 pixels (b). The reference characters are shown for comparison (c).

The same scaling and normalization procedure was performed on the reference characters. Only one reference character for each symbol on number plates was used, except for *D*, *8* and *B*, which were the most difficult to differentiate and were represented by two references. For classification template matching with Euclidean metrics was implemented that was further extended by syntax analysis of Polish licence plates. Naturally with such a simple approach and small learning set it was not possible to achive a near 100% recognition rate, but this was not the aim of the research. The identified number plates were used as unique vehicle identifiers to show the feasibility of our speed measurement method described in Section 3.4.

3.4 Speed Measurement

The proposed system is able to measure the speed of incoming vehicles in real-time (on a fast PC). It uses recognized licence plates to identify approaching vehicles and estimate their velocity. The algorithm is able to cope with more than one licence plate at a time. The principle of speed measurement is the following: if we can measure the distance covered by a vehicle between two consecutive frames, we can calculate its speed. The algorithm localizes all license plates (or the coordinates of their centers of gravity in the binarized image) within both frames and finds ono-to-one correspondence between them. Assuming correct plate recognition on both images, we have – for each observed car – two *image positions*: initial on the first frame and final on the second one. As shown Fig. 4, an *image position y* can be converted to the *road distance d* from the reference point *A*. The difference of the final and the initial *road distance* is the sought lenght of vehicle travel between the two images.

The non-linear relation *d=f(y)* that is used to calculate the *road distance d* based on the *image position y* of an object at the road level was found experimentally by measuring the image position and the road-level position of the

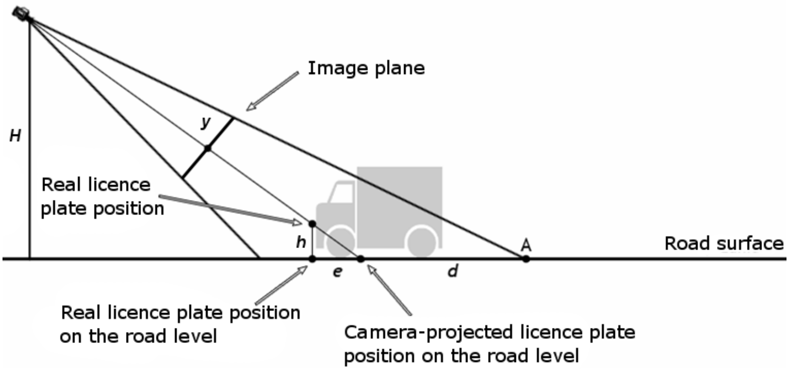


Image plane
y
H
Real licence plate position
h
A
Road surface
e
d
Real licence plate position on the road level
Camera-projected licence plate position on the road level

Fig. 4. The principle of vehicle position measurement with a single camera. The distance d of a vehicle from the reference point A in the camera field of view is proportional to the position of the vehicle in the image plane y.

road stripes on one of the pictures in the sequence shown in Fig. 5. Due to the fact that we use the position of the licence plate center as a reference point, an error e in distance measurement is introduced. This error could be eliminated if the precise elevation over the road-level of a given licence plate h was known. Unfortunately the elevation of licence plates varies. The average height of a licence plate was around 40 cm, while approximately 90% of them were mounted between 30 and 50 cm over the street level (as measured for 100 licence plates on one of the main streets in Warsaw). Simple calculations based on the theorem of Thales yield the following formula for the relative error of the distance covered by a car between two consecutive frames: $\delta = \frac{h}{H-h}$.

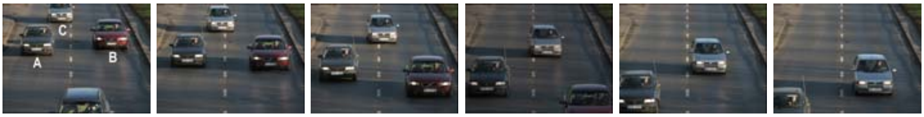


Fig. 5. A typical sequence of images acquired by our system. Car A is being overtaken by cars B and C. It is clear that car A is the slowest and car B is the fastest among the three. Moreover, car A is maintaining a constant velocity of 45±1 km/h as measured by an onboard GPS device. Velocities of cars B and C are unknown. The average measured speed in km/h returned by our system is: 46,0 (A), 83,7 (B), 62,9 (C).

Assuming that the camera is mounted 7.5 meters over the street level (which was the case in our experiments), the relative error of speed measurement for 90% of the vehicles would be between +4.2% and +6.9%. However, if we consider the average elevation of licence plates of 40 cm and assume that for 90% of vehicles the diversion from the average is ±10 cm, the relative error decreases to an acceptable level of ±1.4%. The above holds only if the licence plate is perfectly localized in the image and the *d=f(y)* relation is precisely found. In real-life situations, the imprecision of the these two measurements may contribute to an additional error as shown in Section 4.

4 Experimental Evalution and Results

We evaluated our algorithm[1] on a number of traffic video sequences of type shown in Fig. 5. We used a car with a licence plate elevated exactly 40 cm above the ground and a GPS device for speed measurement. We assume that the driver was able to maintain constant speed with accuracy of ±1 kmh. The collected data for three different velocities are depicted in Table 1 below.

Table 1. Actual car velocities (in km/h) as indicated by GPS versus values measured with the proposed method. The licence plate elevation was precisely known.

Actual velocity	Measured velocities
43 ± 1	43.3; 44.0; 42.9; 44.1; 42.6; 44.0
55 ± 1	54.9; 56.0; 55.9; 56.2; 56.9
71 ± 1	70.7; 71.1; 71.0; 69.1

The measured velocities' errors are practically within the precision range of the GPS. The final error for the proposed method with 90% confidence level at velocities around 100 km/h should not exceed 4% and is less for lower velocities, which is comparable with commerically available speed radars. Unfortunately for rare cases of extremely high mounted licence plates, the error will by much higher and at the moment there is no compensation mechanism implemented although it can be done based on the size of the observed vehicle.

As mentioned in Section 3.3, the character classification algorithm was rather simple and the correct licence plate identification rate was only around 83%.

Image processing time on a laptop (PentiumM 1.8 GHz) computer was around 1 second for each frame (2816 x 2112), so the algorithm can perform in real time on a fast computer without modifications.

5 Conclusions

This paper introduces a novel concept for vehicle speed measurement based exclusively on image analysis. The experimental evaluation shows that the achieved results are very promising and they are on par with comercially available radar based systems. The main advantage of our approach is low cost and high accuracy of the system.

[1] Authors would like thank student Mr. Marcin Lewkowicz for his help in pursuing this research.

References

1. Fan, X., Fan, G.: Graphical Models for Joint Segmentation and Recognition of License Plate Characters. IEEE Signal Processing Letters 16(1), 10–14 (2009)
2. Chen, Z.X., Liu, C.-Y., Chang, F.-L., Wang, G.Y.: Automatic License-Plate Location and Recognition Based on Feature Salience. IEEE Trans. on Vehicular Technology 58(7), 3781–3786 (2009)
3. Iwanowski, M.: Metody morfologiczne w przetwarzaniu obrazow cyfrowych, AOW EXIT (2009)
4. Maduro, C., Batista, K., Batista, J.: Estimating Vehicle Velocity Using Image Profiles on Rectified Images. In: Araujo, H., Mendonça, A.M., Pinho, A.J., Torres, M.I. (eds.) IbPRIA 2009. LNCS, vol. 5524, pp. 64–71. Springer, Heidelberg (2009)
5. Lin, H.Y., Li, K.J., Chang, C.-H.: Vehicle speed detection from a single motion blurred image. Image and Vision Computing 26, 1327–1337 (2006)
6. Donoser, M., Arth, C., Bischof, H.: Detecting, Tracking and Recognizing License Plates. In: Yagi, Y., Kang, S.B., Kweon, I.S., Zha, H. (eds.) ACCV 2007, Part II. LNCS, vol. 4844, pp. 447–456. Springer, Heidelberg (2007)
7. Cho, Y., Rice, J.: Estimating Velocity Fields on a Freeway From Low-Resolution Videos. IEEE Trans. on Intelligent Transportation Systems 7(4), 463–470 (2006)
8. Glowny Urzad Statystyczny, Transport - Wyniki dzialalnosci w 2005 r., Departament Statystyki Uslug GUS, Warszawa (2006) (in Polish)
9. Shapiro, V., Gluchev, G., Dimov, D.: Towards a multinational car license plate recognition system. Machine Vision and Applications 17, 173–183 (2006)
10. Staniak, M., Czajewski, W.: Lokalizacja i identyfikacja obiektow dla celu manipulacji Postepy Robotyki: Systemy i wspoldzialanie robotow. In: Wydawnictwa Komunikacji i Lacznosci, Warszawa, vol. II, pp. 85–94 (2006) (in Polish)
11. Wu, Y., Shen, J., Dai, M.: Traffic object detections and its action analysis. Pattern Recognition Letters 26, 1963–1984 (2005)
12. Iwanowski, M.: Automatic car number plates detection using morphological image processing. Przeglad Elektrotechniczny 81, 58–61 (2005)
13. Chang, S.L., Chen, L.S., Chung, Y.C., Chen, S.-W.: Automatic License Plate Recognition. IEEE Trans.on Intelligent Transportation Systems 5(1), 42–53 (2004)
14. Shapiro, V., Dimov, D., Bonchev, S., Velichkov, V., Gluhchev, G.: Adaptive License Plate Image Extraction. In: Proc. of CompSysTech. 2003 (2003)
15. Hsieh, J.W., Yu, S.-H., Chen, Y.-S.: Morphology-based License Plate Detection from Complex Scenes. In: Proc. of ICPR 2002 Conf. (2002)
16. Dailey, D.J., Cathey, F.W., Pumrin, S.: An algorithm to estimate mean traffic speed using uncalibrated cameras. IEEE Trans. on Intelligent Transportation Systems 1(2), 98–107 (2000)
17. Siyal, M.Y., Fathy, M.: Image Processing Techniques For Real-Time Qualitative Road Traffic Data Analysis. Real-Time Imaging 5, 271–278 (1999)
18. Duric, Z., Rivlin, E., Rosenfeld, A.: Understanding object motion. Image and Vision Computing 16, 185–797 (1998)
19. Broida, T.J., Chellappa, R.: Estimation of Object Motion Parameters from Noisy Images. IEEE Trans. on Pattern Analysis and Machine Intelligence PAMI-8(1), 90–100 (1986)

A Particle-Based Method for Large-Scale Breaking Waves Simulation

Emmanuelle Darles, Benoit Crespin, and Djamchid Ghazanfarpour

XLIM - UMR 6172 - CNRS, University of Limoges
benoit.crespin@xlim.fr

Abstract. We address in this paper the problem of particle-based simulation of breaking waves. We present a new set of equations based on oceanographic research which allow us to deal with several types of breaking waves and multiple wave trains with full control over governing parameters. In order to reduce computations in non-significant areas, we also describe a simple and efficient multiresolution scheme, controlled using the properties of our breaking wave model.

1 Introduction and Previous Works

Early works dealing with the simulation of oceanic scenes used parametric models. Under certain conditions it is possible to obtain realistic shapes of breaking waves [1], but this approach cannot usually represent the intrinsic dynamic nature of the fluid such as interactions and collisions with the shore or other floating objects. A notable exception is the work by Yuksel *et al.* [2]. Another approach consists in solving the dynamic equations governing fluid motion (e.g. Saint-Venant or Navier-Stokes) using an Eulerian method and a static 3D grid. A breaking wave can be represented using this approach by controlling its starting geometry [3]; this approach is suitable both for real-time simulations [4] and computer-generated motion pictures [5]. However non-linear phenomena cannot be represented using this method without a full, 3D resolution [6–8] that can require a very fine-scale grid to capture small details.

Lagrangian methods aim at simulating fluid dynamics using particle systems, for example relying on the Smooth Particles Hydrodynamics (SPH) model [9, 10] that we chose for our implementation. Their main advantage is the ability to cover virtually infinite domains. *Hybrid* methods combine Eulerian grids with a set of particles dedicated to the simulation of small-scale details [11–14]. This approach was used to produce realistic simulations of water pouring into a glass, or splashes and bubbles generated by quantifying the amount of dissipation and turbulences in the case of breaking waves [15, 16]. Another method based on this approach solves Saint-Venant equations governing fluid motion by incorporating a turbulence function generating sprays and foam [17].

The specific problem of breaking waves simulation was addressed in several papers [3, 13, 16, 18]. All of them propose to initialize the velocity field using parametric models in order to obtain a visually convincing break. The basic idea

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 316–323, 2010.

is to "push" the water volume from the rear, by numerically simulating a real wave generator. The main drawback is that only one breaking wave can really be obtained: after a certain simulation time, the magnitude of the velocity field is too low to obtain a new one. As a consequence, simulating *wave trains* is difficult with this approach since re-initializing the velocity field before the fluid returns to rest may imply numerical instabilities. As a consequence, choosing appropriate parameters for the wave generator is an extremely difficult process, making it impossible in practice to simulate successive plunging breakers in the same scene.

In order to improve computation times, especially when dealing with large oceanic scenes, a *multiresolution approach* can be added since non-significant areas of the fluid domain do not require highly precise computations. This idea was developed for Eulerian methods [6, 8] by refining the 3D grid in areas close to the air-water interface or by considering the visual impact on the final rendered image. Lagrangian methods are also well adapted to multiresolution [19–22], since particles' sizes can be fitted according to a "split and merge" scheme that allows to generate more (and relatively smaller) particles in significant areas. However care must be taken to limit the impact of suddenly varying particle size, in order to avoid "popping" effects on consecutive frames.

The approach proposed in this paper aims at representing large scale trains of breaking waves. We show in Sect. 2 how to generate breaking waves that are not limited in time, as well as multiple wave trains. Sect. 3 describes a multiresolution scheme able to generate bigger particles in less significant areas, which reduces aliasing effects observed with previous methods.

2 Simulation of Breaking Waves

2.1 Finite Impulse Method

One of the most used models to describe wave motion in oceanography is the *soliton* [23]. The soliton equation basically represents the propagation of a solitary wave in a non-linear, dispersive medium. It expresses initial horizontal and vertical velocities of a volume of fluid by considering its height H and its depth related to the bottom d. Since it only defines initial velocities, applying the soliton description directly to the SPH solver is not possible because after a while particles don't have enough energy to rise. Therefore, we extend this formulation as an external force called *impulsion*, which will be applied to each particle at each frame. In the 2D case, this force with coordinates v_x and v_y is computed by:

$$\begin{aligned} v_x &= \sqrt{gd}\frac{H}{d} sech^2(A(x - w\bar{t})) \\ v_y &= \sqrt{3gd}\frac{H}{d}^{3/2}\frac{y}{d} sech^2(A(x - wt))\tanh(A(x - w\bar{t})) \end{aligned} \tag{1}$$

The *pulsation* of the soliton, noted w, will be used to obtain different types of breaking waves (see Sect. 2.2). The *soliton time* $\bar{t}$ expresses the time progression of the soliton itself, which is independent of the global simulation time. Other parameters in Eq. 1 are the gravity force g, $A = \sqrt{\frac{3H}{4d^3}}$ and $sech = \frac{1}{\cosh}$.

The horizontal component v_x, always positive, pushes the particle towards the beach. The vertical component v_y pushes or pulls the particle vertically. The impulsion force increases to make the wave rise, then decreases after it reaches its maximum. The soliton can be reinitialized when needed to avoid a temporal limitation. We extend the soliton to the 3D case by adding a *deviation angle* θ from the orthogonal direction to the shore. A new parameter ϕ representing the soliton phase difference is also used to allow a maximal push for any x value:

$$\begin{aligned} \tilde{v_x} &= \sqrt{gd}\tfrac{H}{d} sech^2(A(x - w\bar{t} + \phi)) \\ \tilde{v_y} &= \sqrt{3gd}\tfrac{H}{d}^{3/2}\tfrac{y}{d} sech^2(A(x - w\bar{t} + \phi)) \tanh(A(x - w\bar{t} + \phi)) \\ \tilde{v_z} &= \sqrt{gd}\tfrac{H}{d} sech^2(A(z\sin\theta - w\bar{t} + \phi)) \end{aligned} \quad (2)$$

Component $\tilde{v_z}$ is used pushes particles along direction z in order to obtain asymmetric breaking waves, as can be observed in real ocean scenes, controlled by θ. One important difference with $\tilde{v_x}$ and $\tilde{v_y}$ is that $\tilde{v_z}$ is only applied to breaking particles *i.e.* for which the vertical component $\tilde{v_y}$ decreases, which ensures that no push along z direction occur in stable regions.

The impulsion force f^{sol} with coordinates $(\tilde{v_x}, \tilde{v_y}, \tilde{v_z})$ obtained from Eq. 2 is applied to each particle at each time step as an exterior force; this impulse represents about 10% of the overall forces applied to a particle. Although our method is more physically-inspired than purely physically-based, it is fully controllable as shown in Figure 1 where various heights H and deviation angles θ are specified. To obtain plausible breaking waves, θ should belong to the interval $[O, \pi/4]$.

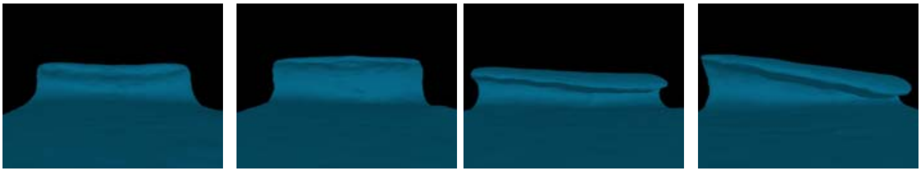


Fig. 1. Different breaking waves with heights ranging from 0.3 to 0.4 and deviation angles from 0 to 0.5

2.2 Controlling Breaking Waves

A useful parameter for describing waves behavior is the *surf similarity parameter* or Iribarren Number N_I [24]:

$$N_I = \frac{\tan\beta}{\sqrt{H/L_0}} \quad (3)$$

where β is the slope angle at the bottom of the sea, H is the waves height and L_0 is the deep water wavelength. A value of N_I greater than 2.0 indicates a surging breaker, a value between 0.4 and 2.0 a plunging wave and a value lower than 0.4 a spilling wave.

We obtain L_0 from Eq. 3 and the characteristic *pulsation* w of the desired type of wave by:

$$L_0 = \frac{N_I^2 H}{\tan^2 \beta} \quad , \quad w = \sqrt{\frac{2\pi g}{L_0}} \tag{4}$$

This value is used in Eq. 2 to compute the soliton equation. Depending on N_I we obtain different types of breaking waves, as shown in Figure 2 with $H = 0.3$ and $\tan \beta = 0.14$.

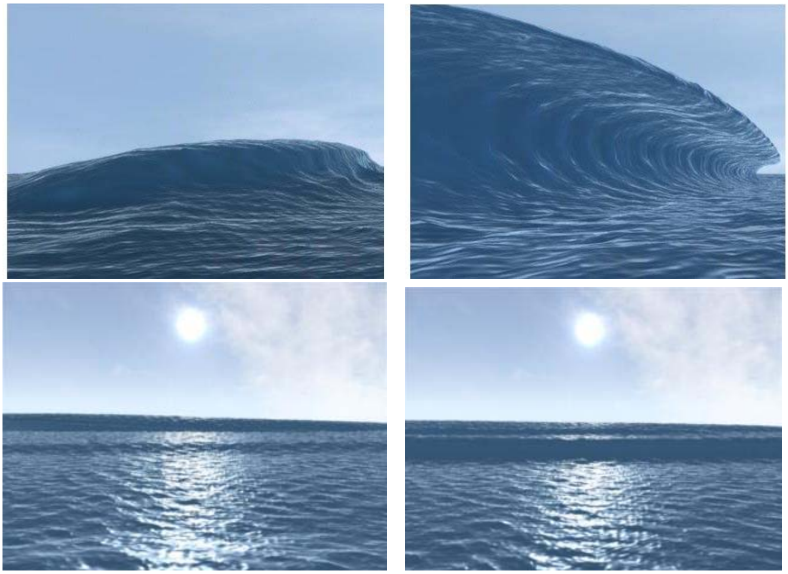


Fig. 2. Top: Breaking waves with $N_I = 0.3$ and 0.5. **Bottom:** Successive breaking waves

2.3 Multiple Impulsions Model

To simulate several waves in the same scene, the total force applied to a particle is given by the sum of N impulsion forces computed for each soliton in the scene. Each force $\boldsymbol{f_k^{sol}}$ $(1 \leq k \leq N)$ is computed by Eq. 2 with its own parameters $\bar{t}_k, H_k$, ϕ_k, θ_k, etc.

Real waves tend to organize themselves as *groups*: successive waves are usually grouped in periods of high waves, followed by periods of low waves. A *wave train* is composed of several waves having similar characteristics and propagating along the same direction. Our method can be used to simulate the behavior of real waves by defining all parameters for a soliton belonging to a wave train in a probabilistic way, allowing the end-user to control the global characteristics of the wave train *and* the relative variation of those characteristics inside the train. Another advantage of this approach is the ability to take into account large-scale phenomena such as tides and currents by choosing similar deviation angles for all waves in the train. An example of two successive waves is shown on Figure 2.

3 Multiresolution Particles

A huge amount of particles are required in order to simulate large ocean scenes, thus inducing high memory and computation costs. Lagrangian multiresolution is a good solution to this problem since it uses a hierarchical representation of the system through two basic operations: merging and splitting. Merging consists in grouping several particles together if they are close enough to each other and share similar properties, and then consider this group as a single bigger particle at further time steps, whereas splitting is the opposite process.

Our approach is based on a *progressive mass transfer*: depending on different criteria, a small fraction of mass can be transferred at each time step either from a particle to its neighbors (i.e. *progressive splitting*), or from its neighbors to a particle (i.e. *progressive merging*, depicted in Figure 3). If a particle has ”lost” all of its mass, it is invalidated, and will not be considered in the simulation anymore. This process guarantees that the overall mass is conserved; our approach also prevents the sudden appearance or disappearance of particles, thus avoiding ”popping” effects.

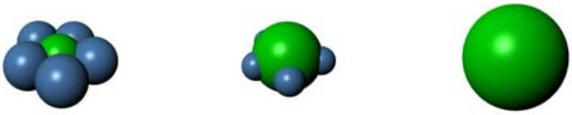


Fig. 3. Progressive merging (from left to right): a fraction of mass is transferred to a particle from its neighbours at successive time steps

Different criteria can be taken into account to decide whether a particle should split or merge with its neighbours. In our case these are:

- the approximate distance from the particle to the free surface (for breaking waves simulation, the inverse distance to the bottom is a good approximation)
- the distance between the particle and the viewpoint
- the impulsion force applied to the particle: if this force has a large value, then the particle should not be merged since it belongs to a breaking wave where intense interactions occur.

Our implementation combines these parameters to ensure that merging particles are those located deep below the surface or near the surface but far from the viewer, with a low visual impact on the final rendering.

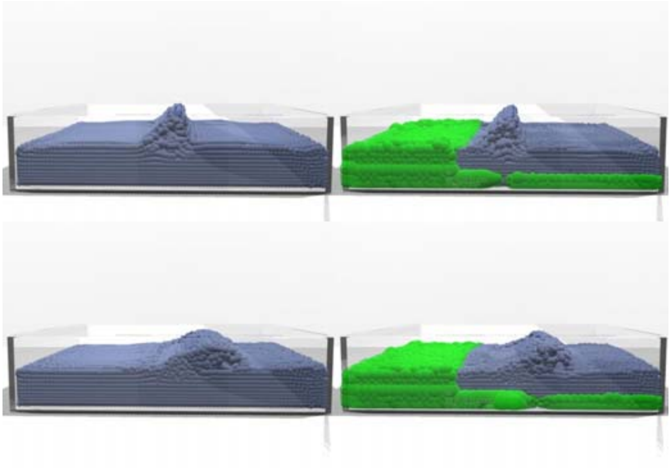


Fig. 4. Comparison between simulations with and without multiresolution, with 100K particles. Green color is used for particles affected by the merging/splitting process.

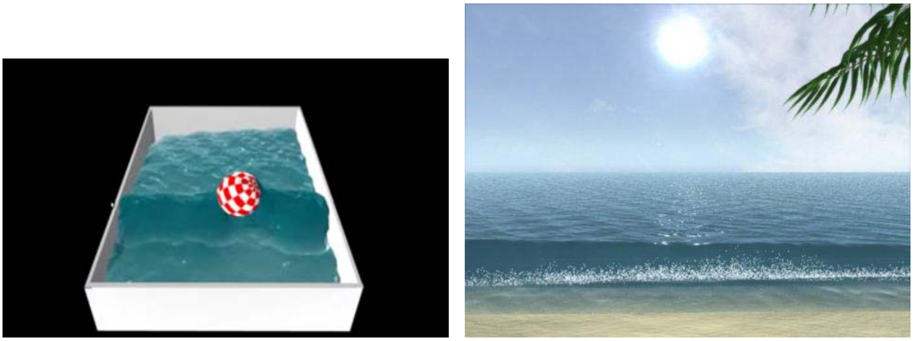


Fig. 5. Interactions of a breaking wave with a ball or with the shore

At each time step, a fraction of mass is transferred between particles; as in previous multiresolution approaches [20], we use the ,,shooting / gathering" method of Desbrun and Cani [19] to efficiently compute pressure and viscosity forces that respect reciprocity between particles of different sizes. The process ends either if a merging particle does not satisfy the aforementioned criteria anymore, a merging particle reaches the maximum mass defined by the user, or a splitting particle has lost all of its mass. In that last case we remove it from the main simulation pool.

Our experiments reveal that using a fraction of 10% of the original mass of the child particles is sufficient to avoid popping effects while being fast enough to reduce computational costs, since we don't have to rely on expensive approximate medial axis computation [20] or static *layers* [19].Our results are illustrated in Figure 4.

Our simulation runs at about 4 sec. per frame with multiresolution enabled, whereas standard simulation runs at 7 sec. per frame on an Intel Core 2 Duo processor running at 2GHz with 2GB of RAM. The gain due to our multiresolution scheme, measured in terms of active particles count, increases linearly. For 200K initial particles, we obtain a gain of 45% less active particles and a total acceleration factor of 1.75, whereas the gain was limited to 10% with the implementation of [20].

4 Conclusion and Future Work

As expected, our results are visually similar to those obtained with 2D methods specifically targeted for breaking waves [3]. However, our method relies on a full 3D simulation, which can thus integrate physically-based phenomena such as collision detection with the shore or other objects (see Figure 5). For rendering purposes we used a simple particle-based spray generator and an animated ocean texture.

Our work takes place in a more general context of realistic ocean scenes rendering, hence future investigations will focus on other non-linear phenomena related to breaking waves such as foam. We also plan to adapt our multiresolution scheme to more general particle-based fluid simulations, and study how it could fit to recent GPU architectures to get better performances.

Acknowledgments

Part of this work was funded by the Dynamé project (French National Research Agency).

References

1. Jeschke, S., Birkholz, H., Shumann, H.: A procedural model for interactive animation of breaking ocean waves. In: WSCG (2003)
2. Yuksel, C., House, D.H., Keyser, J.: Wave particles. In: SIGGRAPH (2007)
3. Milhaef, V., Metaxas, D., Sussman, M.: Animation and control of breaking waves. In: Symposium on Computer Animation, pp. 315–324 (2004)
4. Thuerey, N., Mueller-Fischer, M., Schirm, S., Gross, M.: Real-time breaking waves for shallow water simulations. In: Pacific Graphics (2007)
5. Bredow, R., Schaub, D., Kramer, D., Hausman, M., Dimian, D., Stirling Duguid, R.: Surf's up: the making of an animated documentary. In: SIGGRAPH Courses (2007)
6. Losasso, F., Gibou, F., Fedkiw, R.: Simulating water and smoke with an octree data structure. Computer Graphics, 457–462 (2004)
7. Irving, G., Guendelman, E., Losasso, F., Fedkiw, R.: Efficient simulation of large bodies of water by coupling two and three dimensional techniques. ACM Trans. Graph. 25(3), 805–811 (2006)
8. Losasso, F., Shinar, T., Selle, A., Fedkiw, R.: Multiple interacting liquids. ACM Trans. Graph. 25(3), 812–819 (2006)

9. Monaghan, J.: Smoothed particle hydrodynamics. Annu. Rev. Astron. Physics 30, 543 (1992)
10. Desbrun, M., Cani, M.-P.: Smoothed particles: A new paradigm for animating highly deformable bodies. In: Eurographics Workshop on Computer Animation and Simulation, pp. 62–76 (1996)
11. O'Brien, J., Hodgins, J.: Dynamic simulation of splashing fluids. Computer Animation, 188–205 (1995)
12. Foster, N., Fedkiw, R.: Practical animation of liquids. In: SIGGRAPH (2001)
13. Enright, D., Maschner, S., Fedkiw, R.: Animation and rendering of complex water surfaces. ACM Transaction on Computer Graphics 21 3, 736–744 (2002)
14. Takahashi, T., Fujii, H., Kunimatsu, A., Hiwada, K., Saito, T., Tanaka, K., Ueki, H.: Realistic animation of fluid with splash and foam. Comput. Graph. Forum 22(3), 391–400 (2003)
15. Kim, J., Cha, D., Chang, B., Koo, B., Ihm, I.: Practical animation of turbulent splashing water. In: Symposium on Computer Animation (2006)
16. Losasso, F., Talton, J., Kwatra, N., Fedkiw, R.: Two-way coupled sph and particle level set fluid simulation. Transactions on Visualization and Computer Graphics 14(4) (2007)
17. Thuerey, N., Rude, U., Stamminger, M.: Animation of open water phenomena with coupled shallow water and free surface simulations. In: Symposium on Computer Animation (2006)
18. Becker, M., Teschner, M.: Weakly compressible sph for free surface flows. In: Symposium on Computer Animation, pp. 209–217 (2007)
19. Desbrun, M., Cani, M.-P.: Space-time adaptive simulation of highly deformable substances. Technical Report 3829, INRIA (December 1999)
20. Adams, B., Pauly, M., Keiser, R., Guibas, L.J.: Adaptively sampled particle fluids. In: ACM Transactions on Graphics (SIGGRAPH 2007 papers), vol. 26 (2007)
21. Hong, W., House, D.H., Keyser, J.: Adaptive particles for incompressible fluid simulation. The Visual Computer 24(7-9), 535–543 (2008)
22. Yan, H., Wang, Z., He, J., Chen, X., Wang, C., Peng, Q.: Real-time fluid simulation with adaptive sph. Comput. Animat. Virtual Worlds 20(2&dash;3), 417–426 (2009)
23. Radovitzky, R., Ortiz, M.: Lagrangian finite element analysis of newtonian fluid flows. Int. J. Numer. Meth. Engng. 43, 607–619 (1998)
24. Battjes, J.A.: Surf similarity. In: 14th Coastal Engineering Conference, pp. 466–480 (1974)

A Hierarchical Classification Method for Mammographic Lesions Using Wavelet Transform and Spatial Features

Silvana G. Dellepiane, Irene Minetti, and Sara Dellepiane

University of Genoa, Dept. of Biophysical and Electronic Engineering
Via all'Opera Pia 11a, 16145 Genova, Italy
{silvana.dellepiane;irene.minetti;sara.dellepiane}@unige.it

Abstract. A supervised hierarchical classification for mammographic images is here presented, based on the analysis of wavelet transform. The proposed method analyses different wavelet transform decomposition levels at the same time and innovatively treats them as new images taking into account also their spatial, contextual and morphological properties. The proposed method is very simple and transparent to the user, but it is able to achieve very good classification results. Since it is very fast, we foresee its future application not only for classification purposes, but also for pathology automatic localization. MIAS database has been used in order to compare results with the ones presented in literature.

Keywords: Mammographic CAD, Wavelet Analysis, Morphological Image Analysis.

1 Introduction and Related Work

Despite the quite recent introduction of other diagnostic procedures (i.e., for instance, Magnetic Resonance and Ultrasound), X-ray mammography is still universally employed in early detection of breast cancer. Difficulties in understanding mammograms arise because some pathologies are very small and spatially distributed, image contrast is low and some classes are not precisely separated, tissue mixtures or ambiguous cases are present. For such reasons, automatic Computer-Aided-Diagnosis (CAD) systems may help radiologists in discriminating breast pathologies.

Regarding classification purpose, a large number of papers dealing with the detection of microcalcifications [1]-[3] has been published while papers devoted to the detection and classification of breast masses have more recently been published, even if they still represent a limited number of research works [4]-[6]. At present, only a few papers deal at the same time with both microcalcifications and masses, after having distinguished normal from pathological situations [8]-[9]. In all cases many methods are based on the wavelet approach. When using wavelet transform decomposition, some global sub-band features are usually measured, such as, for instance, energy and entropy [10]-[11], but contextual,

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 324–332, 2010.

morphological and geometrical information are usually lost. Out of a wavelet framework approach, geometrical moments have been seldom applied for the detection of microcalcifications, usually after detecting the region boundaries by means of a segmentation step. Anyway the classical scheme of segmentation and image classification has proved to fail, due to the complexity of the diagnostic images, the undefined borders and shape of normal and pathological tissues, the presence of strongly textured tissues, and the varying and, sometimes, limited size of some lesions. Previous attempts to take geometrical features into account in a multi-resolution wavelet framework include the wavelet modulus-maxima method [12] for mammographic mass shape analysis. As a general statement, most of the cited papers compute a large number of features [9].

The paper presented here addresses the automatic classification of normal and abnormal tissues, separating pathological cases into microcalcifications, circumscribed masses and radial masses. The proposed supervised hierarchical classifier avoids the critical segmentation phase and innovatively exploits the morphological and contextual analysis of the wavelet transform coefficients, by using features related to spatial-scale multi-resolution analysis.

Contrary to other hierarchical approaches, the natural taxonomic decomposition of semantic classes, i.e. normal versus abnormal cases, and microcalcifications versus masses is not followed here. To exploit the feature separability power, the binary tree first-level nodes represent macro-classes that share common geometrical characteristics in the wavelet approximation coefficients. The second tree level uses two different criteria in separating normal from microcalcifications classes and circumscribed from radial masses: microcalcifications are detected through the study of their wavelet details coefficients, that emphasise the presence of intensity changes in the images, such as edges or spots, while masses can be recognised via a deeper analysis of the shape of the brightest wavelet approximation coefficients, that are related with the actual shape of the pathological sign. No pre-processing or equalisation steps are required but features and metrics turn to be adaptive to the single image under analysis. The proposed method is also very simple and transparent to the user, by avoiding the use of a lot of features.

Images taken from MIAS database[7] have been used during the experimental session in order to achieve results comparable with the ones reported in literature. In the following, Section 2 describes in detail the structure of the proposed method, Section 3 presents the experimental session and the obtained results. Finally, conclusions are reported in Section 4.

2 The Proposed Method

Since the purpose of the work is the classification of the pathology after solving the problem of its detection, the original images are cropped, centering a 128x128 window of interest on the abnormality. In [8] a 64x64window is used, but some pathologies, especially some circumscribed masses, are often larger and may be confused with normal light tissue if cropped in such a way. The proposed

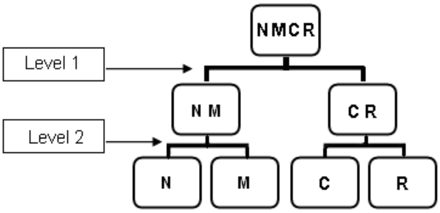


Fig. 1. Classification diagram: N: Normal class; M: Microcalcification; C: Circumscribed mass; R: Radial Mass

classification tree is represented in Fig.1. At the first level, two macro-classes are separated: normal tissue and microcalcifications (NM) versus circumscribed and radial masses (CR). Normal classes and microcalcifications are characterised by almost homogeneous tissue (more or less dense depending on the patient) (Fig. 2a). In this case no compact brighter area, i.e., mass tissue, is concentrated at window centre, as in Fig. 2c and Fig. 2d. Microcalcifications are characterised by a few light spots sparse in the whole window (Fig. 2b) and, even when they are concentrated at the centre of the image, they do not create a large homogeneous area, as masses do. As a consequence, at this level, the classification is based on a preliminary geometrical analysis regarding the lightest parts of the image, to discriminate if a centred mass is present or not.

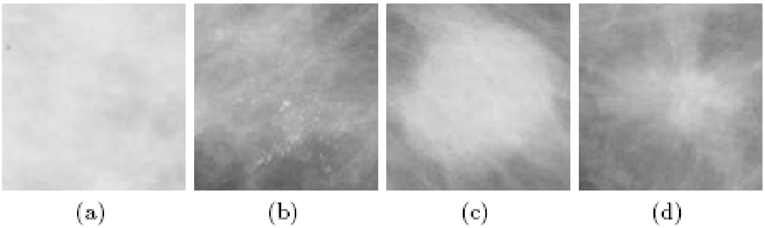


Fig. 2. Examples of breast tissues: (a) normal tissue; (b) microcalcification; (c) circumscribed mass; (d) radial mass

At the second classification level, the analysis is deepened to discriminate each single class, on the basis of its shape details or gradient details. The second level of classification is very differentiated for each macro-class, according to their deep aspect differences. Microcalcifications have to be detected through the study of their wavelet details coefficients, that emphasise the presence of intensity changes in the images, such as edges or spots. Masses can be recognised, instead, via a deeper analysis of the shape of the lightest wavelet approximation coefficients, that correspond to the shape of the present anomaly. In this case, the detail wavelet coefficients are not considered because the images are not characterised by clear edges or other kinds of details.

A training phase is devoted to the evaluation of the optimal thresholds values and to the creation of class templates. In order to avoid the negative influence of mixture classes in this phase, outlier images are automatically located and discarded. The learnt thresholds are then applied during the classification process at the first hierarchical level while the class templates refer to the discrimination within each macro-class.

The following subsections describe in detail the proposed image processing for feature extraction and the classification criteria at level 1 and 2.

2.1 First-Level Classification

Since the spatial arrangements of the brightest pixels in the image is often indicative of the presence of a mass or some microcalcifications, as a first step, each image is decomposed (and downsampled) with a one-level Haar wavelet transform and only the position of the 200 (i.e., the 5%) largest approximation coefficients (LL_1) is considered (Fig. 3). It has been experimentally validated, by preliminary tests, that 200 is the optimal number of coefficients to be considered. The result of this process is a 64x64 binary image, as shown in Fig. 4.

We can empirically verify that the largest coefficients are quite concentrated around the centre of the image if a mass is present, while they are more spread on the whole window in case of normal images (N) and microcalcifications (M).

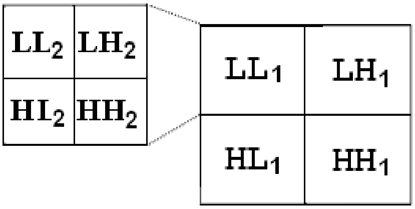
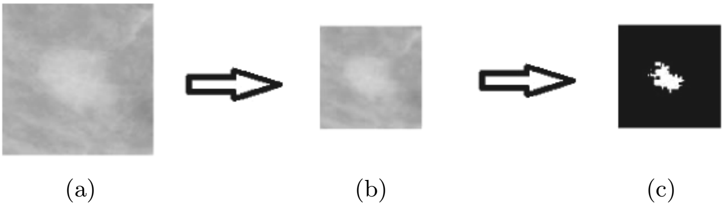


Fig. 3. Wavelet image decomposition at first and second level



(a) (b) (c)

Fig. 4. Selection of the largest approximation coefficients of the image: (a) Original image; (b) One-level approximation coefficients; (c) Mask corresponding to the position of the 200 largest coefficients

As a quantitative measure, able to discriminate between concentrated and sparse objects, the second-order central moment (M_{22}) of the mask is then considered (Eq. 1).

$$M_{22} = \sum\sum(x - x_c)^2(y - y_c)^2 b(x, y) \tag{1}$$

x_c and yc represent the barycentre coordinates and $b(x, y)$ represents the binary value of the mask pixel (x, y). M_{22} offers a measure of the spreading of the coefficients around their barycentre, and its value increases with the spatial dispersion. As a consequence, when masses are present, M_{22} is expected to be much lower than in case of microcalcifications or normal tissues. Another discriminating feature is the distance D of the mask barycentre from the window centre. In fact, by working hypothesis, masses, if present, are expected to be centred in the window, so having a low D value.

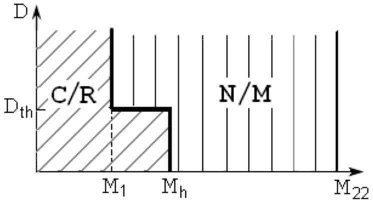


Fig. 5. Discriminating criteria in the (Moment, Distance) feature space

After evaluating M_{22}, the image can be assigned to CR macro-class if $M_{22} < M_l$ or to NM macro-class if $M_{22} > M_h$. If $M_l <_{M22}< M_h$, the parameter D_{th} is calculated and the macro-class is assigned according to the feature partitioning space as in Fig. 5. The values of parameters M_l, M_h, and D_{th} are optimized in the training phase.

2.2 Second-Level Classification

To separate normal from microcalcifications images, the local energy of the wavelet detail coefficients is considered. To this purpose, the original image is decomposed with a two-level Haar wavelet transform and only the detail coefficient matrixes at both levels are considered, i.e. LH_1, HL_1, HH_1, LH_2, HL_2, HH_2. Each matrix is split into 16 identical block sectors and the energy of each sector is calculated. A feature vector of 96 elements (16 elements per matrix) is therefore used for the classification. During the classification step, the Euclidean distance between this feature vector and the mean of the feature vectors for each class (N and M), obtained during the training phase, is calculated and the Minimum Distance to Means (MDM) criterion is then applied.

To discriminate between masses, instead, the 100 largest approximation coefficients of one-level Haar wavelet transform are considered. During the training

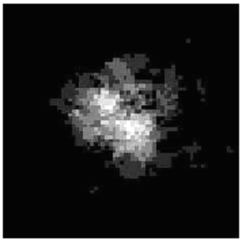


Fig. 6. Example of template for circumscribed masses

phase, a mean mask is computed for each class, thus obtaining a sort of prototype for class C and R.

This process is like the generation of a template but, in this case, it can be considered a fuzzy template since it produces a multi-level and not a binary mask. An example of circumscribed prototype is shown in Fig. 6. During classification, the MDM criterion is applied, considering the 100 lightest approximation coefficient mask of the image under analysis and the templates.

3 Experiments and Results

For a preliminary analysis, the Data Set n.1 has been considered, made up of 25 pathological images and 10 normal images from the MIAS database. It was decided to use half of these data as a training set and the remaining as a test set, but similar performances have been obtained by adopting a leave-one-out training-test approach. 10,000 trials were performed by randomly changing the training and test sets.

Table 1. Confusion matrix obtained during the preliminary analysis

		Output Classification			
		Normal	Microcalcification	Circumscribed	Radial
Input Images	Normal	97%	2%	0%	1%
	Microcalcification	25%	68%	3%	4%
	Circumscribed Mass	0%	0%	81%	19%
	Radial Mass	9%	2%	27%	62%
Overall Accuracy		77%			

The confusion matrix reported in Table 1 points out that normal images and circumscribed masses are classified very well. Radial masses and microcalcifications are more subject to errors. However, an important achieved result is that the error is mostly kept inside each macro-class and not between different macro-classes. This result means that the first level of the hierarchical classifier is very precise, thanks to the appropriate geometrical and morphological image properties taken into account.

Table 1 can also be summarised showing the classification accuracy when considering the distinction between normal and abnormal situations. These values correspond to the classifier sensitivity and specificity reported in Table 2 (first row).

After the preliminary analysis, a larger test session has been performed, by considering 43 pathological images of MIAS database (Data Set n.2), including ambiguous cases, and 18 normal images. Also in this case 10,000 trials have been performed, with different random training sets. With such a data set a few more errors occurred at the first classification level (macro-class identification) but, in general, the classification error is strictly correlated with mixture classes or doubtful situations. For example, some images that were automatically classified differently from MIAS ground truth are reported in Fig. 7, showing the complexity of the task.

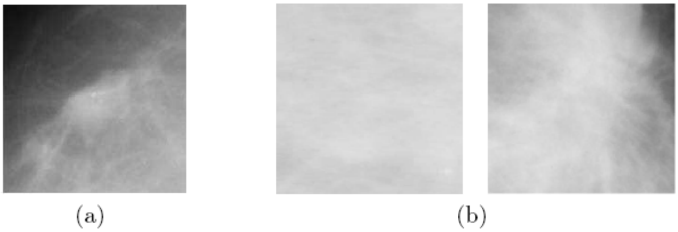


Fig. 7. Examples of misclassification: (a) Circumscribed mass classified as microcalcification; (b) Radial masses classified as normal tissue

Table 2 summarises the results also for the second data set, pointing out sensitivity and specificity in discriminating abnormal images. Performances are satisfying, also when compared to the state of art, like, for example, the system in [9] (last row of Table 2). The confusion matrix of Table 3 concerns the distinction between microcalcifications and masses.

Table 2. Sensitivity and Specificity when discriminating between normal and abnormal cases

Classifier	Specificity (Abnormal)	Sensitivity (Normal)	OA
Proposed Method (Dataset 1)	97%	90%	92%
Proposed Method (Dataset 2)	86%	86%	86%
Method in [9]	72.7%	80%	77.7%

Also in this case, results are comparable to the ones presented in literature: for example, the best result of [9] shows an Overall Accuracy (OA) of 85.4%, while our classifier obtains an OA of 84.15%, with a much simpler approach. In [10] the OA seems to be higher but less images are used as a test.

Table 3. Confusion matrix when discriminating pathological cases (Dataset 2)

		Output Classification	
		Microcalcifications	Masses
Input Images	Microcalcifications	75.6%	24.4%
	Masses	7.3%	92.7%
Overall Accuracy		84.15%	

Table 4. Sensitivity and Specificity when discriminating between masses and not masses

Classifier (Masses vs. Not Masses)	Specificity	Sensitivity	OA	N.Images
Proposed Method (Dataset 1)	83%	94.82%	88.33%	17 masses/18 not masses
Proposed Method (Dataset 2)	86%	82.5%	84.1%	31 masses/30 not masses
Method in [4]	76%	89%	82.5%	45 masses/45 normal
Method in [5]	95%	80%	87.5%	20 masses/20 normal

Finally, Table 4 summarises sensitivity and specificity in discriminating 'masses' versus 'not masses', comparing the promising performances of our method with results proposed by other works. It is worth to notice, however, that in [4] and [5] the discrimination is made between masses and normal images, while our system is able to discriminate masses from normal tissues or microcalcifications.

4 Conclusions

A supervised hierarchical classifier for the automatic classification of normal and abnormal mammary tissue has been presented, able to separate pathological cases into microcalcifications, circumscribed masses and radial masses. Wavelet transform is applied and morphological and contextual analysis is exploited, by using features related to spatial-scale multi-resolution analysis. Results show to be very promising, especially in discriminating normal from abnormal tissue and microcalcifications from masses. All the parameters have been experimentally chosen and good robustness with respect to their changes has been experienced.

References

1. Strickland, N., Hahn, H.I.: Wavelet Transforms for Detecting Microcalcification in Mammograms. IEEE Transactions on Medical Imaging 15, 218–229 (1996)
2. Zhang, W., Yoshida, H., Nishikawa, R.M., Doi, K.: Optimally Weighted Wavelet Transform Based on Supervised Training for Detection of Microcalcifications in Digital Mammograms. Medical Physics 25, 949–956 (1998)
3. Bagci, A.M., Yardimci, Y., Çetin, A.E.: Detection of Microcalcification Clusters in Mammogram Images Using Local Maxima and Adaptive Wavelet Transform Analysis. In: Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing, vol. 4, pp. 3856–3859 (2002)

4. Petrosian, A., Chan, H.-P., Helvie, M.A., Goodsitt, M.M., Adler, D.D.: Computer-Aided Diagnosis in Mammography: Classification of Mass and Normal Tissue by Texture Analysis. Phys. Med. Biol. (39), 2273–2288 (1994)
5. Chiracharit, W., Sun, Y., Kumhom, P., Chamnongthai, K., Babbs, C., Delp, E.J.: Normal Mammogram Classification Based on a Support Vector Machine Utilizing Crossed Distribution Features. In: Proc. of the 26th Int. Conf. IEEE EMBS, vol. 1, pp. 1581–1584 (2004)
6. Zaïane, O.R., Antonie, M.L., Coman, A.: Mammography Classification by an Association Rule-based Classifier. In: Int. Workshop on Multimedia Data Mining, pp. 62–69 (2002)
7. Suckling, J., Parker, J., et al.: The Mammographic Images Analysis Society digital mammogram database. In: Gale, A.G., Astley, S.M., et al. (eds.) Digital Mammography. Exerpta Medica International Congress Series, vol. 1069, pp. 375–378. Elsevier, Amsterdam (1994), http://www.wiau.man.ac.uk/services/MIAS/
8. Ferreira, C.B.R., Borges, D.L.: Analysis of Mammogram Classification Using a Wavelet Transform Decomposition. Pattern Recognition Letters (24), 973–982 (2003)
9. Mousa, R., Munib, Q., Moussa, A.: Breast Cancer Diagnosis System Based on Wavelet Analysis and Fuzzy-Neural. Expert Systems with Applications 28(4), 713–723 (2005)
10. Yang, J.C., Shin, J.W., Park, D.S.: Comparing Study for Detecting Microcalcifications in Digital Mammogram Using Wavelets. In: Yang, Z.R., Yin, H., Everson, R.M. (eds.) IDEAL 2004. LNCS, vol. 3177, pp. 409–415. Springer, Heidelberg (2004)
11. Soltanian, H., Zadeh, F., Rafiee Rad, S., Pourabdollah-Nejad, D.: Comparison of Multiwavelet, Wavelet, Haralick, and Shape Features for Microcalcification Classification in Mammograms. Pattern Recognition (37), 1973–1986 (2004)
12. Bruce, L.M., Adhami, R.R.: Classifying Mammographic Mass Shapes Using Wavelet Transform Modulus-Maxima Method. IEEE Trans. On Medical Imaging 18(12), 1170–1177 (1999)

Easy Rigging of Face by Automatic Registration and Transfer of Skinning Parameters

Ludovic Dutreve[1], Alexandre Meyer[1], Veronica Orvalho[2], and Saïda Bouakaz[1]

[1] Université Lyon 1, LIRIS, CNRS, UMR5205, France
[2] Instituto de Telecomunicações
Faculdade de Cincias da Universidade do Porto, Portugal

Abstract. Preparing a facial mesh to be animated requires a laborious manual rigging process. The rig specifies how the input animation data deforms the surface and allows artists to manipulate a character. We present a method that automatically rigs a facial mesh based on Radial Basis Functions (*RBF*) and linear blend skinning approach. Our approach transfers the skinning parameters (feature points and their envelopes, ie. point-vertex weights), of a reference facial mesh (source) - already rigged - to the chosen facial mesh (target) by computing an automatic registration between the two meshes. There is no need to manually mark the correspondence between the source and target mesh. As a result, inexperienced artists can automatically rig facial meshes and start right away animating their 3D characters, driven for instance by motion capture data.

1 Introduction and Related Work

Modeling 3D faces is becoming more and more automatic and common with systems based on photos [1] or on user-friendly interactions [2]. Consequently, many applications in the area of games or virtual reality offer to novices the capacity to generate or customize a 3D facial avatar. Indeed, avatar with better ability to depict his owner in the virtual community will provide a better immersion. However, setup a facial mesh in order to animate it requires fastidious manual rigging to specify how the input animation data deforms the surface. For instance, the first stage of a rigging process, which is to manually place some feature points (FP) on the face, takes around 2 minutes [3]. Thus, in the same way that [4] does for the body, we aspire to a system that automates the face rigging task, to make animations more accessible for children, educators, researchers, and other non-expert animators.

For the real-time facial mesh deformation, we have chosen to focus on the traditional Linear Blend Skinning[1] (LBS) [5,6]. Indeed, with Blendshape [7] which consists on morph between key meshes, they remain the most popular methods used in practice in real-time 3D engines. Although both methods provide realistic animations, it seemed more tractable for our purpose to deal with skinning parameters than key shapes, and more pertinent to stay compatible with recent

[1] Also known as skinning or enveloping or Skeletal Subspace Deformation(SSD).

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 333–341, 2010.

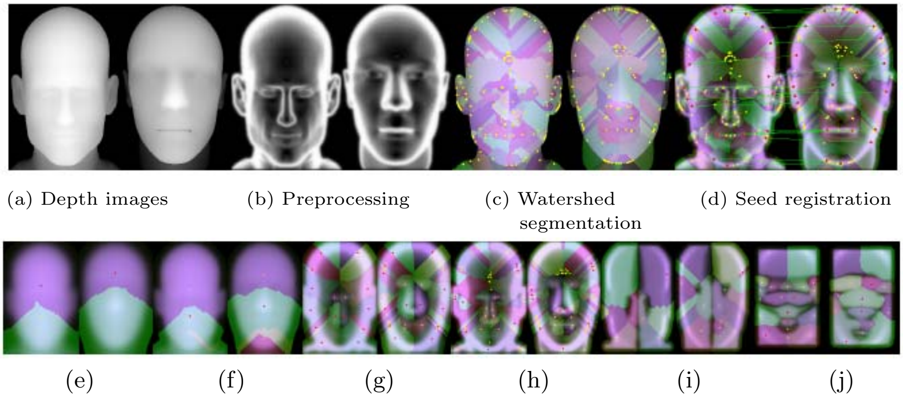
mesh deformation based on skinning [8] for a potential extension. Skinning binds controllers to vertices in order to deform the mesh according to the controllers deformations. Usually, these controllers are bones or skeleton for the body and feature points (FP) or landmarks for the face[2]. In most case, a vertex may be influenced by several FP by using a convex combination of weights.

In this paper, we propose a method which takes as input a mesh of face (target) and provides as output skinning parameters, *i.e.* a FP set and the associated point-vertex influences (weights). These skinning parameters are transferred without human intervention from a reference mesh (source) already rigged. Comparing to *procedural* approaches mostly used in automatic body rigging approaches like [4], transfer-based approaches has the advantage, in the specific case of face, to reuse fine manual work designed specially to tackle subtle facial animation. Moreover, previous transfer approach for face [10] has validated the transfer based concept. Comparing to [10], our main contribution is the fully automatic registration which provides a landmark correspondence between the two meshes (See Section 2). Once the 3D target mesh is rigged with our method, it may be directly animated by any FP-based animations coming for instance from a motion capture system based on webcam as illustrated in our results (See Figure 4). Or, it may be used by any 3D engines and major 3D Softwares.

An important aspect of the automatic transfer of facial animation parameters is the registration which tries to find a transformation that optimally fits points from the source surface to the target surface. Registration of range images of the same object [11] is often tackled by variant of the Iterative Closest Points algorithm [12,13,14]. Although we represent the frontal face view with a depth image which is similar to a range image, our problem has to be classified in the area of registering two instances of objects belonging to the same family: variation of faces morphologies. Our case is also more general than the class of non-rigid registrations problems aiming at fitting a template [15,7,16,14] model to scanned or motion capture surface. These methods are often based on variational approaches minimizing surfaces differences from morphable shape model computed from a database.They provide good results when faces are close which is not necessary our case. And, for the sake of generality, we preferred working on an approach without *a priori* on the surfaces which allows to register two cartoon faces as illustrated in Figure 2d and 2e.

Overview. A reference facial mesh is previously rigged for skinning animation. Our method automatically registers the reference mesh (source) to the facial mesh to rig (target) in order to transfer the skinning parameters (FP and point-vertex influences) without human intervention. The result is a rigged facial mesh ready to be animated by skinning. The transfer is done in four phases. Firstly, the method we present addresses the correspondence issue using a new approach registering the result of a sequence of watershed segmentations described in Section 2. Then, these landmarks are used in a radial basis functions (*RBF*) learning process to transfer the FP from the source to the target and in the

[2] As the ones defined by the Facial Definition Parameters of the MPEG4 standard [9].


(a) Depth images (b) Preprocessing (c) Watershed segmentation (d) Seed registration

(e) (f) (g) (h) (i) (j)

Fig. 1. Top row: one step of our landmark based registration. Bottom row: a sequence of the watershed segmentation/registration process with different preprocessing. (e) and (f): Gaussian filters with different radius. (g) and (h): a Laplacian filter followed by a Gaussian filter with different radius. (i) and (j): vertical and horizontal anti-symmetric filters.

third phase (Section 3) to computes a dense registration. And finally, the point-vertex influences are transferred.

2 Surface Correspondence

The first issue of our transfer approach is the computation of the landmark correspondence on both meshes, which may be tricky if the anthropometric lengths are slightly different. To find this landmark correspondence, our approach runs a sequence of preprocessing, watershed segmentation and region-based registration on the frontal depth images of the faces.

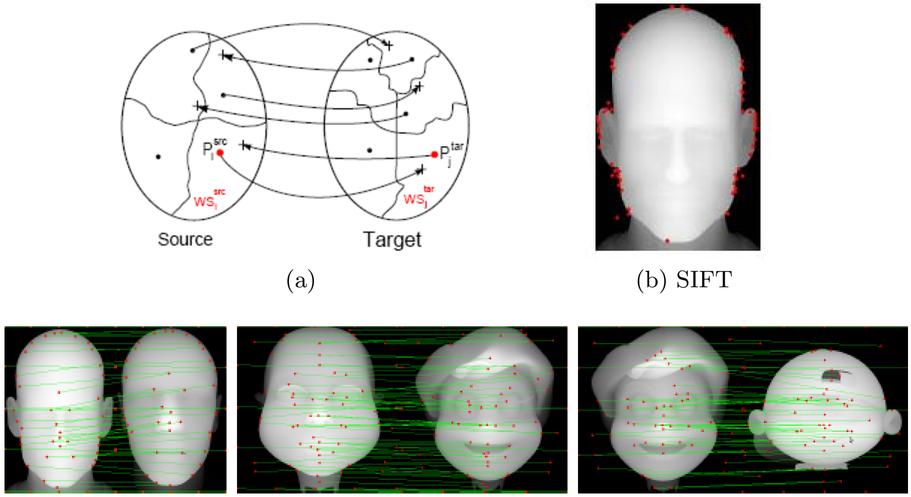
We have chosen to register the two faces using a depth images representation for several reasons. We start from the observation that a frontal part of a face is 2.5D, so the depth image representation should keep the important information for registration. Depth image allows to be easily independent on the mesh resolution which is essential for the usefulness of the method. A low polygon facial model may be registered with a high polygon model. Comparing to complex remeshing, computing the depth map representation of a face can be done easily using an orthogonal projection (as for instance the Z-buffer) and the reprojection of a 2D pixel position on the mesh allows to return in 3D. And finally, working on regular 2D grids (images) is more convenient and more efficient in computation time than working directly with the mesh.

One Step of Watershed Segmentation. At each step, to compute new landmark correspondences, we perform a watershed segmentation where the regions are registered. The depth image is segmented by the watershed algorithm [17] interpreting it as a topographic surface and gradually immerses it in a water container [17]. After the segmentation, the surface will be divided into its catchment basins meaning every pixels is assigned to a region as illustrated in Figure 1c. The local minimum point (seed) of each basin is the representant of the basin. Since the segmentations are computed using the global image information, our method can localize the landmark correspondence more accurately than local detection/grouping methods [18]. For instance, the SIFT descriptor algorithm selects landmarks only on the border which is less precise than our method as illustrated in Figure 2b and 2c. The assumption of our method is that objects of a same family will provide mostly similar segmentation which are more easy to register than directly the meshes.

At step 0, to initialize the sequence, we set landmarks around the respective bounding box of each face. In this paragraph, we assume that k steps of the sequence of watershed segmentation/registration have already been done and have provided a set of landmarks. These landmarks are used as learning points for a RBF interpolation as described in [19] defining the transformation $RBF_k^{src->tar}$, just noted RBF, which transforms a point of the source image to the target image. By inverting the landmarks, we compute $RBF_k^{tar->src}$ noted RBF^{-1}. We consider the next step of the sequence, the $k+1^{th}$ watershed segmentation/registration. The watershed segmentation computes on the two depth images, two sets of regions with their associated seeds (See Figure 1c). A seed of the source segmentation will be selected as landmark if it can be associated with a seed of the target segmentation. The correspondence criteria between two watershed is described in Figure 2a. Two watersheds WS_i^{src} and WS_j^{tar} match if the seed P_i^{src} (resp. P_i^{tar}) falls into the watershed WS_j^{tar} (resp. WS_i^{src}) after being transformed by the RBF (resp. RBF^{-1}) defined by the previous steps. *i.e.* if $RBF(P_i^{src}) \in WS_j^{tar}$ and $RBF^{-1}(P_j^{tar}) \in WS_i^{src}$ we add P_i^{src} and P_j^{tar} to the set of landmarks defining the correspondence. For instance, in Figure 1, the bottom right watershed regions match, top ones not. At the end of this step, we recompute the RBF with the added landmarks for the next step. This criteria is fast to compute and has the advantage to quickly registers similar regions. According to our tests, it is enough efficient to not considering an other criteria more complicated to compute.

Sequence of Segmentation/Registration as a Descriptor Problem. The watershed segmentation applied on depth image combined with our registration criteria has the advantage to provide a global approach which can be easily iterated with different preprocessing. Indeed, in this Section we propose to take advantage of various descriptors to catch different regions of a face, for instance regions of high curvature or of high symmetry, *etc.*. Nevertheless, this general framework may be refined by any kind of surface descriptors and is related to the classical issue of choosing pertinent descriptors for image or surface registration [20,11]. The choose of descriptors order have to be from coarse to fine.

The first preprocessing we apply is a Gaussian filter on the depth images to remove the issue of small regions during the segmentation. This will register the global appearance (local maxima) of the face, mainly the nose and often the forehead (See Figure 1e). In second, we apply a Laplacian filter followed by a Gaussian filter with different radius. It generally registers edges like borders of the face or of the nose (See Figure 1g). Finally, since a face included many symmetric regions, we compute two descriptors of symmetry, horizontal and vertical, as illustrated in Figure 1i In practice, results of Figure 3 have been computed by these preprocessing also illustrated in the bottom row of Figure 1. But, reader may define other interesting descriptors for different class of shape.



(a) (b) SIFT

(c) (d) (e)

Fig. 2. In (a), the criteria to match two segmented regions. In (b), the SIFT descriptors algorithm [21] selects landmarks only on the border. In (c), (d) and (e), our sequence of segmentations/registrations provides a set of 2D pair of landmarks. (d) and (e) illustrate the registration on cartoon faces.

3 Transfer of Feature Points and Point-Vertex Influences

The sequence of watershed segmentations/registrations previously described provides as output two sets of landmarks on the mesh defining a correspondence as illustrated in the bottom row of Figure 2. We use these landmarks to define a RBF interpolation $RBF_{final}^{src->tar}$ to transfer each FP position of the source face to the target face as illustrated in the top row of Figure 3.

Rigging the target mesh with common skinning techniques requires to attach each vertex to one or more FP with influence values. As for the FP, we perform this task by transferring this information from the source mesh by defining a dense registration in a similar way than in [22,10]. To compute this dense registration, each vertex V of the target mesh is transformed by the RBF interpolation: $V' = RBF_{final}^{tar->src}(V)$. Since V' does not fall necessary exactly on

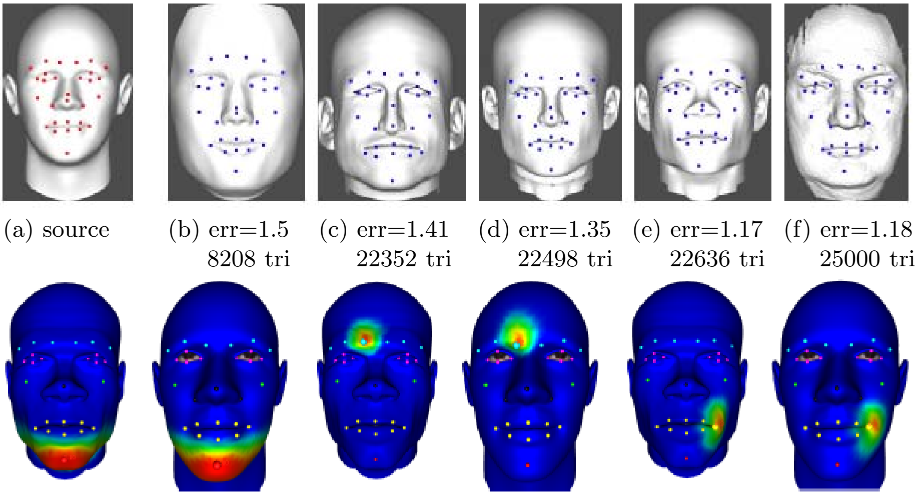

(a) source (b) err=1.5 8208 tri (c) err=1.41 22352 tri (d) err=1.35 22498 tri (e) err=1.17 22636 tri (f) err=1.18 25000 tri

Fig. 3. Top row: examples of transfer of the set of FP. Bottom row: examples of transfer of the point-vertex influences from the reference face (the color scale of the weights starts from blue (0.) to red (1.)).

the source mesh, it is projected[3] on the source mesh to falls into a point P of a triangle T_{src} which defines our dense registration. Notice that P is a 3D position on the triangle and is not necessary a vertex. We compute the FP influences of V by interpolating the ones of the 3 vertices of the triangle T_{src}. We have tested the Inverse Distance Weighting interpolation and the barycentric interpolation which both provide similar and enough accurate results in our case. The bottom row of the Figure 3 shows an example of influences transfer for some FP.

4 Results and Conclusion

Figure 3 shows a sample of our facial mesh database used for our experiments which includes 15 face variations with a large scale of characteristics: 3D meshes count from about thousand triangles to several tens of thousand triangles; bounding box sizes are also quite variable; some meshes include only the face, others include the whole head; and some models are computer generated, others are scanned faces (like the last on the right). The only constraint we ask to the models is that the face has to be manually set to look in the Z axis direction to be able to compute the depth image of the frontal view. We also tested our registration on cartoon faces as illustrated in Figure 2d and 2e. To measure the quality of the landmark correspondence, we have manually defined twenty FP on representative positions for animation for each 3D mesh (See Figure 3a). We

[3] We apply a spherical projection but our tests with a cylindrical projection do not lead to significant differences.

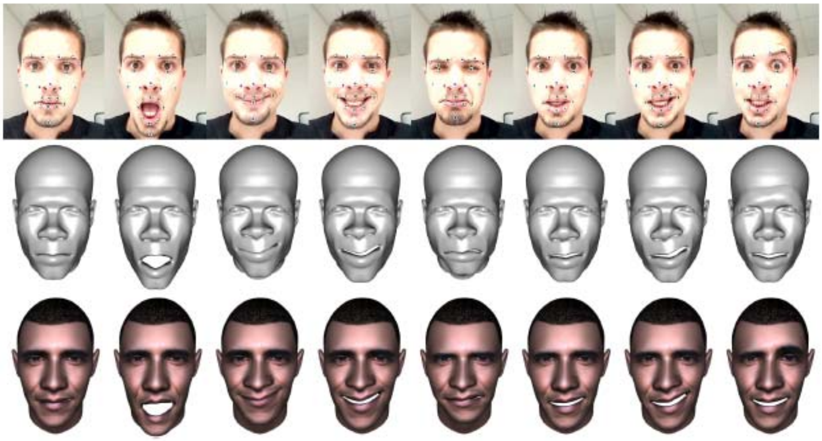


Fig. 4. Top: motion capture animation. Middle: reference model previously rigged. Bottom: the same animation applied on a face automatically rigged by our method, rigging is transferred from the reference without intervention.

define an error criteria by summing the distance between each FP of the transferred rig and the manually set ones normalized by the diagonal of the bounding box to be able to compare results on different mesh size. Notice that the result of the automatic FP transfer may be manually edited if the user feels the need, without changing the rest of the method. Our results in Figure 3, 4 and on the provided video are not manually adjusted.

Computing the depth image representation of the mesh is instantaneous using the GPU/Z-buffer of the frontal view. The landmark correspondence determination is computationally dependant on the depth image resolution and on the number of steps in the sequence of segmentations/registrations. In our experiments, we used depth images of 512 × 512 with the sequence of preprocessings/segmentations/registrations described in Figure 1 and in Section 2. It takes about 20s on a Athlon X2 3800+ with 2GB of RAM. The dense registration is computationally dependent on the number of triangles of the two meshes and takes for instance about 30s for a mesh with 12000 triangles. Once the dense registration is done, the time of skinning parameters transfer is immediate. And finally to illustrate our results, Figure 4 and the provided video show an animation applied on two faces, the first was rigged manually while the second was automatically rigged by our method in a minute without human intervention. The animation is provided on the fly by a motion capture system using a simple webcam and the Lucas-Kanade [23] marker tracker. Notice that any other motion capture system may be used and that the markers on the face may be different than the FPusing retargeting adaptation [24,19].

These results prove the viability of such an automatic transfer approach based on depth image representation of the mesh which makes the implementation easier, lets use efficiently several descriptors with an approach based on segmentation, and allows to easily deal with mesh of different resolutions. Nevertheless, fully automatic transfer of rigging from a generic reference human face to a large range of faces like cartoons or animals is still delicate because of the landmark correspondence issue. A solution would be to rig several types of reference faces (human, cartoon, animal, *etc.*) as preprocess and rig a new face by transfer of parameters from the nearest reference face.

References

1. Geoegios Stylianou and Andreas Lanitis. Image based 3d face reconstruction: a survey. International Journal of Image and Graphics (2009)
2. Ersotelos, N., Dong, F.: Building highly realistic facial modeling and animation: a survey. The Visual Computer 24(1), 13–30 (2007)
3. Softimage. Pick the landmarks with face robot (2009), http://facerobot.wiki.softimage.com/index.php/Stage_3:_Pick_the_Landmarks
4. Baran, I., Popović, J.: Automatic rigging and animation of 3d characters. ACM Trans. Graph. 26(3), 72 (2007)
5. Magnenat-Thalmann, N., Laperrire, R., Thalmann, D.: Joint-dependent local deformations for hand animation and object grasping. In: Proceedings on Graphics Interface (1988)
6. Lewis, J.P., Cordner, M., Fong, N.: Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. In: ACM SIGGRAPH 2000 (2000)
7. Deng, Z., Noh, J.Y.: Computer Facial Animation: A Survey. In: Neumann, U., Deng, Z. (eds.), ch. 1. Springer, Heidelberg (2007)
8. Feng, W., Kim, B., Yu, Y.: Real-time data driven deformation using kernel canonical correlation analysis. In: ACM SIGGRAPH (2008)
9. Zhong, J.L.: Flexible face animation using mpeg-4/snhc parameter streams. In: IEEE International Conference on Image Processing, pp.II: 924–928 (1998)
10. Costa Orvalho, V., Zacur, E., Susin, A.: Transferring the rig and animations from a character to different face models. Computer Graphics Forum (2008)
11. Salvi, J., Matabosch, C., Fofi, D., Forest, J.: A review of recent range image registration methods with accuracy evaluation. Image Vision Comput. (2007)
12. Besl, P.J., McKay, N.D.: A method for registration of 3-d shapes. IEEE Trans. Pattern Anal. Mach. Intell. 14(2), 239–256 (1992)
13. Rusinkiewicz, S., Levoy, M.: Efficient variants of the ICP algorithm. In: Third Int. Conf. on 3D Digital Imaging and Modeling (3DIM) (June 2001)
14. Eisert, P., Schneider, D.: Algorithms for automatic and robust registration of 3d head scans. Journal of Virtual Reality and Broadcasting 2009 (2009)
15. Allen, B., Curless, B., Popović, Z.: The space of human body shapes: reconstruction and parameterization from range scans. In: ACM SIGGRAPH (2003)
16. Weise, T., Li, H., Van Gool, L., Pauly, M.: Face/off: live facial puppetry. In: Symposium on Computer Animation (2009)
17. Mangan, A.P., Whitaker, R.T.: Partitioning 3d surface meshes using watershed segmentation. IEEE Trans. on Visu. and Computer Graphics (1999)
18. Arora Varsha Hedau, H., Ahuja, N.: Matching images under unstable segmentation. In: IEEE Computer Vision and Pattern Recognition (2008)

19. Dutreve, L., Meyer, A., Bouakaz, S.: Feature Points Based Facial Animation Retargeting. In: ACM Symp. on Virtual Reality Software and Technology (2008)
20. Zitova, B., Flusser, J.: Image registration methods: a survey. Image and Vision Computing (2003)
21. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision (2004)
22. Noh, J.Y., Neumann, U.: Expression cloning. In: ACM SIGGRAPH (2001)
23. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Int. Conference on Artificial Intelligence (1981)
24. Pighin, F., Lewis, J.P.: Facial motion retargeting. In: ACM SIGGRAPH 2006 Courses, p. 2. ACM, New York (2006)

Terrain Modeling with Multifractional Brownian Motion and Self-regulating Processes

Antoine Echelard[1], Jacques Lévy Véhel[1], and Olivier Barrière[2]

[1] Regularity Team, Inria, ECP, Grande Voie des Vignes Chatenay-Malabry, France
[2] Faculté de Pharmacie Université de Montréal Pavillon Jean-Coutu, Local 3204
2940 Chemin de la polytechnique Montréal
{antoine.echelard,olivier.barriere,jacques.levy-vehel}@inria.fr

Abstract. Approximate scale-invariance and local regularity properties of natural terrains suggest that they can be a accurately modeled with random processes which are locally fractal. Current models for terrain modeling include fractional and multifractional Brownian motion. Though these processes have proved useful, they miss an important feature of real terrains: typically, the local regularity of a mountain at a given point is strongly correlated with the height of this point. For instance, young mountains are such that high altitude regions are often more irregular than low altitude ones. We detail in this work the construction of a stochastic process called the Self-Regulated Multifractional Process, whose regularity at each point is, almost surely, a deterministic function of the amplitude. This property makes such a process a versatile and powerful model for real terrains. We demonstrate its use with numerical experiments on several types of mountains.

Keywords: Digital elevation models, Hölderian regularity, (multifractional) Brownian motion.

1 Motivation and Background

A digital elevation model (DEM) is a digital representation of ground surface topography or terrain. DEM are widely used for geographic information systems and for obtaining relief maps. In most places, the surface of the earth is rough and rapidly varying. 2D random processes are thus often used as models for DEM. In addition, most terrains possess a form of approximate self-similarity, i.e. the same features are statistically observed at various resolutions. For these reasons, fractal stochastic processes are popular models for DEM.

The most widely used set of models is based on fractional brownian motion (fBm) and its extensions. fBm has been used for mountain synthesis as well as in the fine description of the sea floor [11]. One of the reasons for the success of fBm is that it shares an important property of many natural grounds: statistically, a natural ground is the same at several resolutions. fBm allows to model this scale-invariance property as well as to control the general appearance of the ground via a parameter H taking values in $(0, 1)$: H close to 0 translates into

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 342–351, 2010.

an irregular terrain, while H close to 1 yields smooth surfaces. More precisely, one can show that, almost surely, the local regularity of the paths of fBm, as measured by the Hölder exponent (see definition 2), is at any point equal to H. A large body of works has been devoted to the synthesis and estimation of fBm from numerical data.

A major drawback of fBm is that its regularity is the same at every point. This does not fit into reality: for example, erosion phenomena will smooth parts of a mountain more than others. Multifractional Brownian motion (mBm) goes beyond fBm by allowing H to vary in space. Multifractional Brownian motion simply replaces the real parameter H of fBm with a function, still ranging in $(0, 1)$, with the property that, at each point (x, y), the Hölder exponent of a realization is $H(x, y)$. mBm therefore makes it possible to model a terrain which is smooth at parts and irregular at others, or to synthesize an region containing both a plain and a mountain.

While mBm does improve on fBm for DEM modelling, it does not allow to account in a simple way for the following natural fact: typically, altitude and regularity, in a given terrain, are correlated. For instance, in young mountains, regions at low altitudes are generally smoother than regions at high altitudes: indeed, we have evidenced empirically a link between regularity and altitude on various mountainous zones (see below). Moreover, the sign of the observed correlation allows, to a certain extent, to classify mountains according to their erosion degree. This seems to indicate that mountains should be modeled through a new kind of process, where amplitude and regularity are linked, rather than as an mBm, where the regularity is tuned independently from the amplitude. The purpose if this work is to introduce the Self-Regulating Multifractional Process (SRMP) as a means to do so.

This remainder of this paper is structured as follows: section 2 describes in detail the technical aspects of fBm, mBm and SRMP and includes simulations that allow one to form an idea of the behaviour of these processes. Section 3 presents a method for estimating the parameters. We present in section 4 the results of the modeling by an SRMP of mountain images that were obtained from the *United States Geological Survey* site. Finally, section 5 recapitulates the obtained results and gives prospects for future research.

2 Several Models for Natural Terrains

2.1 Fractional Brownian Motion

Definition and First Properties. Fractional brownian motion was introduced by Kolmogorov in 1940 [7] and further studied by Mandelbrot and Van Ness in [9]. It has proved a useful model in many applications, including financial mathematics, synthesis of artificial mountains or submarine floors [11] or the study of Internet traffic [8].

fBm is a Gaussian centered process with stationary increments. It is parametrized by a constant H. It is self-similar, which means that, denoting the process

$B_H(t)$, $B_H(\alpha t)$ follows the same law as $\alpha^H B_H(t)$ for every fixed $\alpha > 0$. A formal definition is as follows:

Definition 1. *The bidimensional isotropic fractional Brownian motion with Hurst parameter H is the centered Gaussian process B_H such that, for every $(x, y) \in \mathbb{R}^2$:*

$$\mathbb{E}\left(B_H(\boldsymbol{x})B_H(\boldsymbol{y})\right) \propto \left\|\boldsymbol{x}\right\|^{2H} + \left\|\boldsymbol{y}\right\|^{2H} - \left\|\boldsymbol{x}-\boldsymbol{y}\right\|^{2H}, \tag{1}$$

where $||.||$ denotes the Euclidean norm. Fractional Brownian motion is an extension of the ordinary Brownian motion: for $H = 1/2$ it reduces to the well-known Wiener process.

Regularity Properties. fBm is a continuous but almost surely non-differentiable process. Its regularity can be more precisely measured with the help of the pointwise Hölder exponent, whose definition we recall now in a particular case that will be of interest for us.

Definition 2. *Let $\alpha \in (0, 1)$. Let f be a function from $\mathbb{R}$ to $\mathbb{R}$ and $x_0 \in \mathbb{R}$.*

$f \in C^\alpha(x_0)$ if there exists $\epsilon > 0$ and $C \in \mathbb{R}$ such that $\forall t$, $|t| < \epsilon$: $|f(x_0 + t) - f(x_0)| \leq C\,|t|^\alpha$.

The pointwise Hölder exponent of f at x_0 is: $\alpha_p(f, x_0) = \sup\{\alpha \in (0, 1) : f \in C^\alpha(x_0)\}$.

This exponent may be roughly understood as follows: the largest increments $f(x) - f(x_0)$ in the neighbourhood of x_0 are of the order of $|x - x_0|^{\alpha_p(f,x_0)}$.

The parameter H of fractional Brownian motion allows to control its regularity in the sense of Hölder exponent. Indeed, almost surely, for all t, $\alpha_{B_H}(t) = H$. A a consequence, the higher H, the smoother are the trajectories. Figure 1 illustrates this remark.

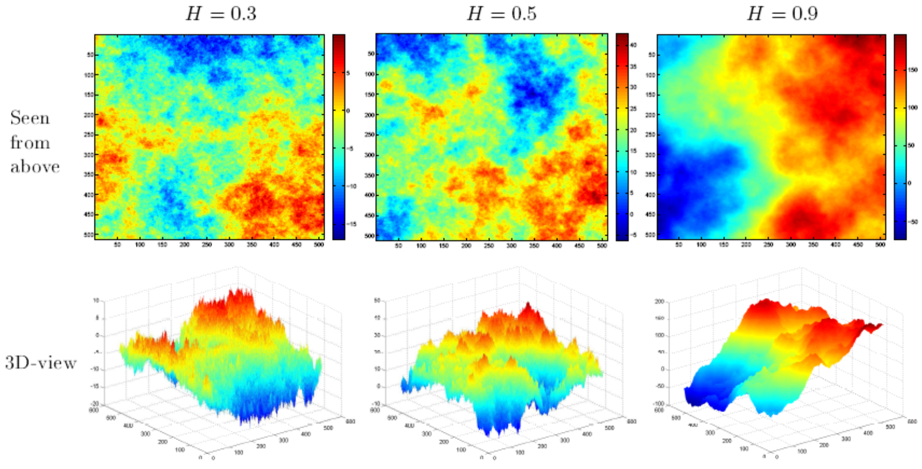


Fig. 1. Realisations of a fractional brownian motion with Hurst parameter H

2.2 Multifractionnal Brownian Motion

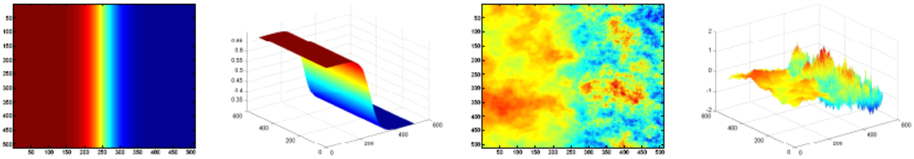
Motivation and Definition. As mentioned above, an important drawback of fBm for terrain modeling is its constant regularity. Indeed, erosion and meteorological phenomena, for instance, smooth some parts of real terrains more than others. More generally, it is obvious that the earth surface presents a variable regularity. This evolving regularity has important consequences in various applications and should be taken into account by a proper model. Multifractional brownian motion, introduced in [10,4], allows to describe such a behaviour.

Definition 3. *Let H be a continuous function from $\mathbb{R}^2$ to $\mathbb{R}$. The multifractional Brownian motion is the centered Gaussian process with covariance function:*

$$\mathbb{E}\left(B_{H(.)}(\boldsymbol{x})B_{H(.)}(\boldsymbol{y})\right) \propto \|\boldsymbol{x}\|^{H(\boldsymbol{x})+H(\boldsymbol{y})} + \|\boldsymbol{y}\|^{H(\boldsymbol{x})+H(\boldsymbol{y})} - \|\boldsymbol{x}-\boldsymbol{y}\|^{H(\boldsymbol{x})+H(\boldsymbol{y})} \quad (2)$$

Regularity Properties. As in the case of the fractional Brownian motion, we use the Hölder exponent to measure the regularity of the trajectories. Assume that H is a differentiable function. Then one may show that almost surely, for all $(x,y) : \alpha_{B_{H(x,y)}}(x,y) = H(x,y)$.

Figure 2 illustrates the variable regularity property of mBm, with a trajectory obtained for a logistic H function. mBm is a quite versatile model. It allows to deal with numerous phenomena, at the expense of more mathematical complexity. As explained in the previous section, a further generalization is needed in order to model images where the regularity depends on the amplitude of the process, rather than being tuned exogenously.

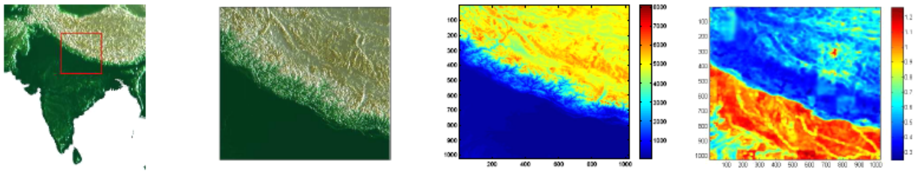


a) b) c) d)

Fig. 2. A realisation of a 2D-mBm with functional parameter H. a) Logistic function H (seen from above); b) Logistic function H ()3D-view); c) mBm with parameter H (seen from above); d) mBm with parameter H (3D-view).

2.3 Self-regulating Multifractional Process

The motivation for introducing self-regulating processes is based on the following general fact: "in nature, the local regularity of a phenomenon is often related to its amplitude". In addition to natural terrains, we have verified this fact experimentally on temperatures records and on RR intervals extracted from ECG [3,5]. Figure 3 exemplifies this on a part of Himalaya.



Himalaya Sub-image Matlab view Hölder exponents

Fig. 3. Zones of high altitude in Himalaya have smaller regularity than regions of lower altitude

A self-regulating process is by definition a stochastic process such that, almost surely, at each point, the Hölder exponent is a deterministic function of the amplitude. In other words, almost surely, at every point x, $\alpha_p(x) = g(Z(x))$ where α_p is the Hölder exponent, Z the process (e.g., the altitude) and g a deterministic function characterizing the process. There are many ways to design self-regulating processes. A natural one is to modify mBm so as to obtain the desired property [3]. This process is called the self-regulating multifractional process (SRMP). As figure 4 shows, the SRMP allows to synthesize realistic terrains. On these landscapes, the "plain" areas (with low altitude) are "automatically" smoother than "mountainous" areas (with high altitude) through a specific choice of the function g.

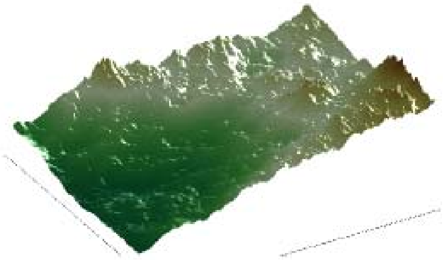


Fig. 4. A realisation of a bidimensional self-regulated multifractional process with $g(Z) = (1 - Z)^2$

Construction. Technically, the SRMP is build using a bidimensional Gaussian field of fractional Brownian motions (introduced in [2]). This field, denoted $\{B(H, x, y)\}_{(H,x,y)\in[a,b]\times[0,1]^2}$ is defined on $[a, b] \times [0, 1]^2$ where $[a, b] \subset (0, 1)$.

For any non-constant continuous f function defined on a compact set $K \subset \mathbb{R}^2$ and ranging in $\mathbb{R}$, and any two reals α et β, we denote: $\overline{\underline{f}}_{\alpha}^{\beta} = \alpha + (\beta - \alpha)\frac{f - \inf_K(f)}{\sup_K(f) - \inf_K(f)}$.

Definition 4. *Let g be a k-Lipschitz function defined on $[\alpha, \beta]$ with values: $[a, b] \subset (0, 1)$.*

Let α' and β' be two random variables such that $\alpha \leq \alpha' < \beta' \leq \beta$. The stochastic operator $\Lambda_{\alpha',\beta'}$ from $\mathcal{C}([0,1]^2, [\alpha,\beta])$ into itself is defined almost surely as:

$\Lambda_{\alpha',\beta'}(Z) = \overline{\underline{B_{g(Z)}}}_{\alpha'}^{\beta'}$ *where $B_{g(Z)}$ is the function $(x,y) \mapsto B_{g(Z(x,y))}(x,y)$.*

Proposition 1. *Almost surely, $\Lambda_{\alpha',\beta'}$ has a unique fixed-point, that is, there exists exactly one function $(x,y) \to Z^*(x,y)$ such that: $Z^* = \overline{\underline{B_{g(Z^*)}}}_{\alpha'}^{\beta'}$*

Definition 5. *The self-regulating multifractional bidimensional process (SRMP2D) parametrized by the function g is defined almost surely as the only fixed-point Z^* of $\Lambda_{\alpha',\beta'}$.*

Regularity Properties of the SRMP. The following important property is shown in [3]:

Theorem 1. *Let Z_g be an SRMP2D. Almost surely, $\forall (x,y) \in [0,1]^2$: $\alpha_{Z_g}(x,y) = g(Z_g(x,y))$*

In order to use the SRMP2D as a model for real terrains, it is necessary to estimate the function g from numerical data. This is the topic of the next section.

3 Evaluation of the Hölder Exponent

The method for estimating the local regularity is based on the so-called generalized quadratic variations [6]. More precisely, we have used the technique presented in [1], which we briefly recall. Consider a discretized mBm $\tilde{B}$ parameterized by a function H with values in $[a,b] \subset (0,1)$. Choose δ, γ such that $\delta - \gamma > \frac{1}{4}$ and $\gamma > \delta b$.

Definition 6. $\tilde{V}_N(t) = \sum_{p \in \tilde{\nu}_N(t)} \left(\sum_{k \in F} d_k \tilde{B}\left(\frac{p+k}{N^\delta}\right)\right)^2$ *where $t = (t_1, t_2)$ and $\tilde{\nu}_N(t) = \tilde{\nu}_N^1(t_1) \times \tilde{\nu}_N^2(t_2)$ andfor $i = 1, 2$:*
$\tilde{\nu}_N^i(t_i) = \left\{ p_i \in \mathbb{N}; 0 \leq p_i \leq N-2 \text{ and } \left| t_i - \frac{p_i}{N^\delta} \right| \leq N^{-\gamma} \right\}$.

Theorem 2. *Almost surely :* $\lim_{N \to \infty} \frac{1}{2\delta}\left(2(1-\gamma) - \frac{\log \tilde{V}_N(t)}{\log N}\right) = H(t)$.

This theorem gives an algorithm to estimate the exponent of an mBm that is discretized at locations $t_{ij} = (i/N, j/N), i,j = 0, ..., N-1$ (we denote $\tilde{B}(t_{ij})$ its values): one first computes $\tilde{V}_N(t_{ij})$ for every t_{ij} and then evaluate its exponent at point t_{ij} with: $\hat{H}(t_{ij}) = \frac{1}{2\delta}\left((1-\gamma) - \frac{\log \tilde{V}_N(t_{ij})}{\log N}\right)$.

This method suffers from the following drawback. Assume $\tilde{B}$ is multiplied by a constant G. Then $\tilde{V}_N(.)$ is multiplied by G^2 and an offset appears: $\Delta_G = H_{GB} - H_B = -\frac{\log G}{\delta \log N}$.

In practice, N has a finite value and the offset of the estimator $\frac{\log G}{\log N}$ can be non-negligible. It is possible to design techniques that correct this offset [3], but we will not use them here.

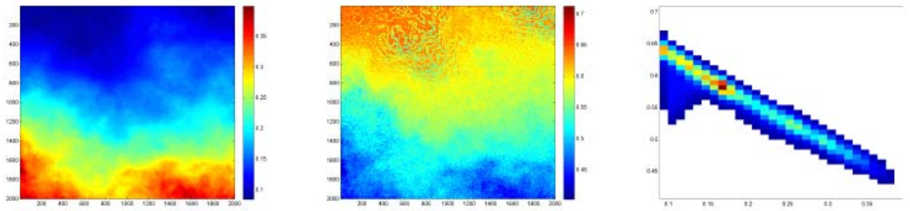
Indeed, what we are looking for is mainly a relation between the exponent and the altitude. The existence of such relation is not affected by this offset. One must however keep in mind that, in the following, the exponents are evaluated within an additive constant.

Theorem 2 shows that this method is theoretically justified in the case of an mBm: if the resolution is good enough then the estimator will be close to the true Hölder exponent.

We now turn to our problem, *i.e.* checking for the existence of a relation g between exponents and altitude on images. To estimate this function g, we proceed as follows:

1. The pointwise Hölder exponent is evaluated at each point of the image. We use the method described above under the assumption that the signal is an SRMP. In this case, contrarily to the situation of mBm, it has not been proved that it gives the correct result. However, numerical experiments on simulated SRMP have shown that it seems to behave well.
2. We draw a scatter plot in the "phase plane", that is the altitude-Hölder exponent plane: each point (x, y) of the image with altitude $z(x, y)$ and estimated regularity exponent $h(x, y)$ is drawn as a point with coordinates $(z(x, y), h(x, y))$ in this plane.

If a relation $h = g(z)$ exists, then every point in the phase plane should approximately belong to the graph of the function g. In order to improve the readability of the scatter plot, a plot of the histogram is drawn instead: the plane (altitude, exponent) is divided into squares whose color becomes "hotter" as more points belong to the square. To test the method, we applied it to an actual SRMP defined by the function $g(Z) = 1 - Z$. As one can see it on figure 3, the points obtained are, within an offset, very close to the graph of the function. This suggests that the previous scheme, although crude, allows to detect relationships between exponents and altitudes.



a) b) c)

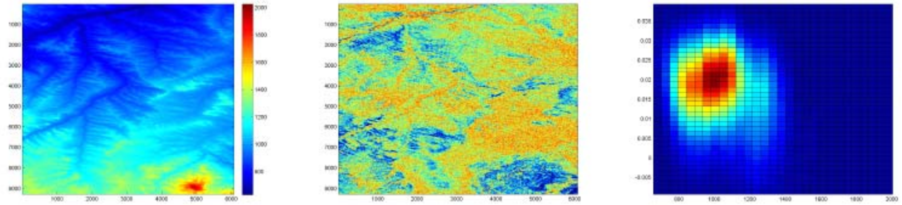
Fig. 5. The SRMP and its evaluated exponents are related: a) A realization of the SRMP b)Its exponent at each point c) Density of the points in the "exponent-altitude" plane

4 Study of Mountain Images

4.1 Study of Global Images

We have studied four mountain images. Two of them are ”young” mountains: Himalaya and Rocky Mountains. The other two are ”old” mountains: Massif Central (France) and Tibesti (Chad). The data were obtained from the internet site of the United States Geological Survey. Their resolution is 3 arc seconds (approx. 90 m). They have the following dimensions: Tibesti: 6097x6393; Massif Central: 3807x3835; Rocky Mountains: 9274x6072; Himalaya: 5620x6767.

In order to check for a possible relation of the type $\alpha = g(z)$, we represent as explained above the scatter diagram in the ”altitude-exponent” plane. However, the results are not satisfying. Indeed, very scattered exponents with the same altitude are observed. As a consequence, the points can not fit the graph of a function. Rather, roughly elliptic “blobs” are observed. See figure 6. This tends to indicate that very large grounds are not homogeneous enough to be represented by an SRMP. In order to find such a relation, we now study subsets of the images.



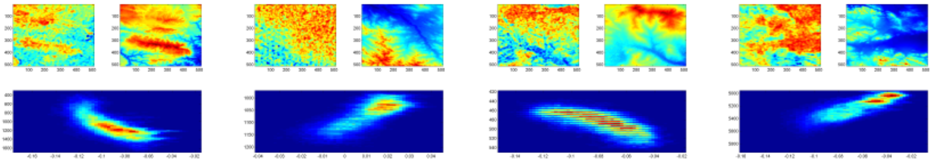
a) b) c)

Fig. 6. Rocky Mountains: a) Original image b) Estimated exponents c) Density of the scatter plot in the “exponent-altitude” plane

4.2 Research on Sub-images

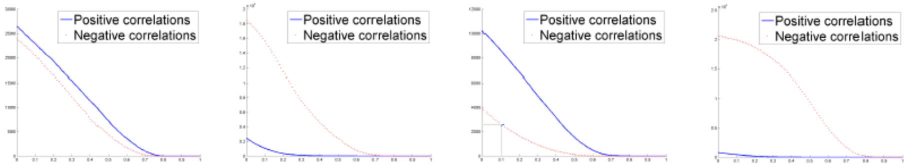
A correlation coefficient with a high absolute value shows the existence of an affine relationship that is verified in a approximate way. For this reason, we have been looking for sub-images of the original images on which the correlation coefficient between the Hölder exponent and the altitude is high. In that view, a square window of size 256^2 or 512^2 pixels was moved over each of the four previous images by steps of 50 pixels. The correlation between the Hölder exponent and the altitude was computed on each sub-image.

On many windows, the correlation is significant. For images with side of length 256 pixels, the percentage of correlation coefficients with higher absolute value that 0.6 are 3.6%, 23.4%, 10.7% et 3.9% respectively for Tibesti, Himalaya, Massif Central and Rocky Mountains. For images with side of length 512, these proportions are respectively 1.7%, 16.6%, 8.9% and 1.7%. Figure 7 gives an example for each mountain and window size 512. The results were similar with size 256.



Massif central Rocky mountains Tibesti Himalaya

Fig. 7. In each cell, the upper-right figure is the original sub-image of length 512 pixels, the upper-left figure displays the exponent at each point of the sub-image and the lower figure shows the density of the scatter plot in the phase plane



Massif central Rocky Moutains Tibesti Himalaya

Fig. 8. For a given abscissa x, in red, the number of windows with a correlation higher than x and in blue, the number of windows with a correlation smaller than $-x$. For instance, the blue circle on the Tibesti figure means: "2600 windows have a correlation smaller than -0.1"

Moreover, on both young mountains, negative correlations are much more frequent than positive correlations. This means that higher peaks are craggier. To quantify this observation, we plot on figure 8 the following graphs: in red, the number of windows with a correlation higher than x and in blue, the number of windows with a correlation that is smaller than $-x$ (this is for 256^2 images. Again, similar results are obtained for 512^2 ones). The absence of significant positive correlations between exponent and altitude appears to be characteristic of young mountains. Indeed, on both young mountains, the few positive correlations are not significant: less than 0.4. The situation is reversed for older mountains, probably because of erosion: high-altitude surfaces of old mountains are smoothed out by erosion, and thus high altitudes translates into more smoothness. For Tibesti, positive correlations are predominant. For Massif Central, the situation is balanced : approximately as many positive correlations as negative ones are observed.

5 Conclusion and Perspectives

We have been looking for a relation between Hölder exponent and altitude. This does not seem to occur for large pieces of land, but significant correlations are observed on smaller parts of terrains. This suggests the existence of a linear relation between exponents and altitude.

Moreover, negative correlations are largely predominant in sub-images from young-mountain images whereas positive significant correlations occur on old mountains. Indeed, the peaks of young mountains are craggier than plains whereas high-altitude surfaces of old mountains are smoothed out by erosion. This property allows to distinguish between young and old mountains.

Because of the rough way of estimating the exponent, we have only been looking for linear relationships. In the future, it would be interesting to estimate more finely both this relationship and the regularity exponents. This would allow to model in a more satisfying way natural terrains as realizations of SRMP.

References

1. Ayache, A., Lévy-Véhel, J.: Identification of the pointwise holder exponent of generalized multifractional brownian motion. Stochastic Processes and their Applications 111, 119–156 (2004)
2. Ayache, A., Taqqu, M.S.: Multifractional processes with random exponent. Stochastic Processes and their Applications 111(1), 119–156 (2004)
3. Barrière, O.: Synthèse et estimation de mouvements browniens multifractionnaires et autres processus à régularité prescrite. Définition du processus autorégulé multifractionnaire et applications. PhD thesis, IRCCyN (2007)
4. Benassi, A., Jaffard, S., Roux, D.: Elliptic Gaussian random processes. Rev. Mat. Iberoamericana 13(1), 19–90 (1997)
5. Barrière, O., Lévy-Véhel, J.: Local regularity analysis of rr intervals. In: CBMS 07 (2008)
6. Istas, J., Lang, G.: Quadratic variations and estimation of the local hölder index of a gaussian process. Ann. Inst. H Poincaré 4(33), 407–436 (1997)
7. Kolmogorov, A.N.: Wienersche spiralen und einige andere interessante kurven im hilbertschen raume. Doklady 26, 115–118 (1940)
8. Mannersalo, P., Norros, I.: Multifractal analysis of real ATM traffic: a first look (1997)
9. Mandelbrot, B., Van-Ness, J.W.: Fractional brownian motion, fractional noises and applications. SIAM Review, 422–437 (1968)
10. Peltier, R.F., Lévy-Véhel, J.: Multifractal brownian motion: definition and preliminary results. Technical report, INRIA (1995)
11. Pesquet-Popescu, B.: Modélisation bidimensionnelle de processus non stationnaires et application à l'étude du fond sous-marin. PhD thesis, Laboratoire d'Electricite Signaux et Robotique, ENS de Cachan (1998)

Image Encryption through Using Chaotic Function and Graph

Sa'ed Faridnia[1] and Karim Fae'z[2]

[1] Islamic Azad University, Qazvin Branch, Qazvin, Iran
[2] Amirkabir University of Technology, Tehran, Iran
faridnia@qazviniau.ac.ir, Kfaez@aut.ac.ir

Abstract. In this research paper, we proposed a new method for image encryption through using chaotic function and graph theory. In this technique, we used a graph for making the coding algorithm more complicated with higher security and changing the gray value of the original image. Experimental results denote that the method benefit from high efficiency against prevalent attacks; for example, the obtained entropy value is 7.9911 that is very close to ideal value of 8.

Keyword: Image encryption, graph, chaotic function.

1 Introduction

Nowadays, the fast development of multimedia products and extensive promotion of digital products on internet has increased the importance of protecting digital information from being illegally copied and distributed more than before. To achieve this goal, various algorithms have been proposed for image encryption [2-5]. Considering the recent great developments in using chaotic function in various applications, many researchers have focused on applying the chaotic functions to image encryption [6-10]. Some of most important advantages of chaotic functions are their high sensitivity to initial values and their noise-like behavior, while being absolutely determined. In [6], a method has been proposed for image encryption through using changing the values of pixels in spatial domain. In [7], a chaotic key-based algorithm (CKBA2) has been proposed for image encryption. In this method, a chaotic function has been employed for changing gray value of image pixels, the following researches illustrated that this method alone is not secure enough[8].

In this paper, a new method has been proposed for image encryption through using chaotic function and graph theory, for creating more complexity and higher security. Using graph theory causes no access to the real values of the gray level of each pixel even if the initial value of chaotic function is detected. In the following sections, we firstly explain the chaotic function; then we present the proposed method. In the section allocated to experimental results, the efficiency of the method against attacks is evaluated by displaying various images.

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 352–359, 2010.

2 Chaotic Function

Chaotic function has a noise-like appearance but is absolutely determined. It means that if we have the initial values and the chaotic map function, we can regenerate the same values. The advantages of chaotic function are:

a) Sensitivity to the initial value
It means that a very small variation in initial value causes an evident difference in the following values of function. In order words, when the initial conditions change a little, the function value changes a lot.

b) Random-like behavior
In comparison with the general random number generators in which the string of generated random numbers can't be regenerated, the methods employed for generating random numbers in algorithms based on chaotic function make the regeneration of the same generated numbers possible, of course, in the case that we have the initial value and the map function.

c) Deterministic Action
Although chaotic functions behave random-like but they are absolutely determined. It means that if we have the initial values and map function, then, we can generate a set of numbers that are apparently random and also regenerate them. Equation (1) shows one of the most popular chaotic functions which is well known as the Logistic Map.

$$X_{n+1} = rX_n(1 - X_n) \tag{1}$$

The Logistic Map function with initial values of $X_0 = 0.5$ and $r = 3.9999$ has chaotic behavior completely. Figure 1 shows the chaotic behavior of the function.

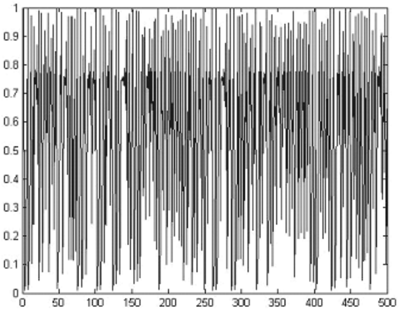


Fig. 1. The chaotic behavior of equation function (1) in first 500 iteration

3 Proposed Method

In this method, we generate a set of numbers between 0 and 255 which are non-repetitive and random in order through using Logistic Map function in order to be inserted in the nodes of a graph. The Logistic Map function used for generating the numbers needs an initial value. For making the proposed method more secure, we use a key with the length of 80 bits for generating the initial value (see equation 1). This key can be defined in ASCII as below (equation 2):

$$K = K_0, K_1, \ldots, K_9(ASCII) \tag{2}$$

That K_i specifies a 8 bits block from key. The key transforms to binary form (equation 3).

$$K = \begin{pmatrix} K_{01}, K_{02}, K_{03}, K_{04}, K_{05}, K_{06}, K_{07} \\ K_{08}, \ldots, \ldots, \ldots, \ldots, K_{91}, K_{92}, K_{93} \\ K_{94}, K_{95}, K_{96}, K_{97}, K_{98}, (Binary) \end{pmatrix} \tag{3}$$

The initial value obtained from equation (4).

$$X_0 = \begin{pmatrix} K_{01} \times 2^{79} + K_{02} \times 2^{78} + \\ K_{11} \times 2^{71} + K_{12} \times 2^{70} + \\ \ldots, \ldots, \ldots, \ldots, \ldots, \ldots \\ K_{n7} \times 2^{1} + K_{n8} \times 2^{0} \end{pmatrix} \tag{4}$$

In the other hand, as shown in figure 1, the function changes between 0 and 1. Therefore, we divide this area into P sections, the size of each is obtained from equation (5).

$$\varepsilon = 1/P \tag{5}$$

Therefore, the area of i'th section is:

$$((i-1)\varepsilon, i\varepsilon) \tag{6}$$

Where P is 256, that is the number of grey level.

In the next stage, we calculate the first value of the function (X1) and specify which section includes this amount. We choose the number of this slice as the first order, providing that the function has not already been located in this domain. This action iterates until hits all of the P areas. Finally, there is a set of non-repetitive numbers in rage of (0,255) as follows:

$$Iteraction = (it_1, it_2, \ldots, it_r) \tag{7}$$

Now, the first number of the iteration order is considered as the first node value in our graph. Then, the second number of the iteration order stands as the second node value in the graph and this node joins the first one. The same procedure resumes until the last node joins to the first one. So we have the circular graph. For example if the order of radon and non-repetitive numbers in iteration order is as ***Iteration =(17,183,45,...8)***, then, the resulting graph is as below:

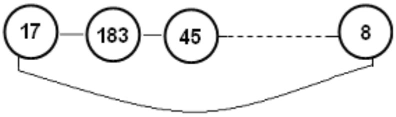


Fig. 2. The resulting graph with iteration order numbers

We have a graph with 256 nodes in which each node contains a non-repetitive number between 0 and 255.we use this graph to determine the gray values of the image pixels. In this stage, the image pixels lie in a row as they set consecutively in a one-dimensional matrix as below (the matrix of the image is saved as lines:

$$V = \{v_1, v_2, \ldots\ldots, v_{M\times N}\} \tag{8}$$

Where, M and N are number of rows and columns. Now for changing the gray value of each pixel, we act as below:

a) For each pixel, the gray value of the pixel is found in an assigned graph with chaotic function and that node is called i.
b) The relative position of the pixel in V is determined and called j, (the index of each pixel in V is its position in the order).
c) According to the size of j, we proceed forward from i in the graph and then we reach to a node, then, the gray value of the destination node is XORed with gray value of the image pixel. The result is inserted as a new gray value of the image pixel.

This method is used for all pixels in V.

4 Experimental Results

A proper encryption method should be robust and secure against all kind of attacks such as cryptanalytic attacks, statistical attacks and brute-force attacks. In this section, we analyze the proposed method of statistical analysis, sensitivity analysis and analysis of key space.

The experimental results show that the proposed method is robust and secure against different attacks.

4.1 Histogram Analysis

The histogram displays the number of pixels in any gray value for any image. The Fig. 3(a) shows the original image and Fig. 3(b), Fig. 3(c) and Fig. 3(d) illustrate the histograms in red, green and blue levels respectively. Also the Fig. 3(e) shows the encrypted image (with ABCDEF0123456789ABCD key in 16-scale basis) of the original image (fig. 3(a)) and fig. 3(f), fig. 3(g) and fig. 3(h) demonstrate the histograms in red, green and blue levels respectively.

As is clearly seen in fig. 3, the histograms of encrypted image are uniform diagrams and these are entirely different from the histogram diagrams of the original image. This makes statistical attacks more difficult.

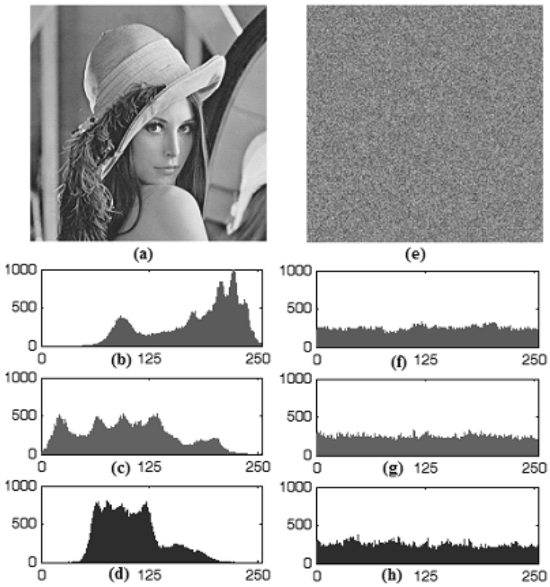


Fig. 3. (a) The original image, and (b), (c) and (d) respectively show the histogram of the lena image with the size of 256 × 256 in red, green and blue level, and (e) illustrates the encrypted image through using the key, ABCDEF0123456789ABCD in a 16-scale. (f), (g) and (h) show the histogram of the encrypted image in red, green and blue level.

4.2 Correlation Coefficients Analysis

In this section, we will consider the correlation of pixels in horizontal, vertical and diagonal directions. For this purpose, 4096 random pairs of neighboring pixels are considered in horizontal, vertical and diagonal directions as the sample. Gray level distribution of neighboring pixels in the original image and encrypted one can be seen in fig. 4. Fig. 4(a), fig. 4(b) and fig. 4(c) show the gray level distribution of horizontal, vertical and diagonal neighboring pixels in the original image respectively and fig. 4(d), fig. 4(e) and fig. 4(f) illustrate the gray level distribution of horizontal, vertical and diagonal neighboring pixels in encrypted image respectively.

Moreover, the correlation coefficient of two neighboring pixels is calculated by using the formula (12).

$$E(x) = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{9}$$

$$D(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i - E(x_i)) \tag{10}$$

$$cov(x, y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - E(x_i))(y_i - E(y_i)) \tag{11}$$

$$r_{xy} = \frac{cov(x, y)}{\sqrt{D(x)\sqrt{D(y)}}} \tag{12}$$

That $E(x)$ is the estimation of mathematical expectations of x, $D(x)$ is the estimation of variance of x, and $cov(x, y)$ is the estimation of covariance between x and y, where x and y are grey-scale values of two adjacent pixels in the image and r is the correlation coefficient. In Table 1, the obtained correlation coefficient values are found.

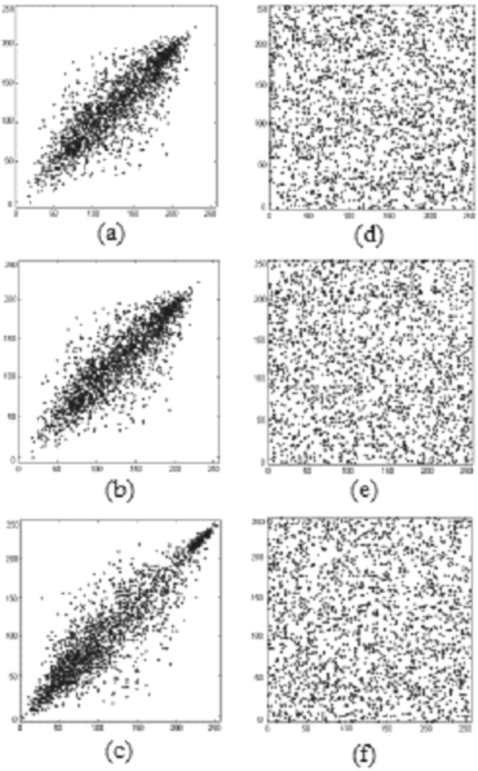


Fig. 4. Gray level distribution of neighboring pixels in (a) horizontal, (b) vertical and (c) diagonal directions of the original image and gray level distribution of the neighboring pixels in (d) horizontal, (e) vertical and (f) diagonal directions of the encrypted image

Table 1. The correlation coefficient values of two neighboring pixels in horizontal, vertical and diagonal directions in the original and encrypted images

Direction	Original image	Encrypted image
Horizontal	0.9432	0.0289
Vertical	0.9027	0.0191
Diagonal	0.8748	0.0163

4.3 Key Space Analysis

In a proper cryptography method, the key space should be extended sufficiently so that it will be resistant to brut-force attacks. In our proposed method, there can be $2^{80}(\approx 1.20893 \times 10^{24})$ different combinations of the key. The experimental results indicate that the different combinations are sufficient for resisting to all types of brute-force attacks.

4.4 Information Entropy

Entropy is one of the main properties for displaying randomness. Information entropy is the mathematical theory of data Communication and storage developed in 1949 by Claude E. Shannon[14]. One of the most famous formulas to calculate entropy is:

$$H(S) = \sum_{i=0}^{2N-1} P(s_i) \log(\frac{1}{P(s_i)}) \tag{13}$$

Where N is the number of gray levels used in the image (in 8-bit images it will be 256) and $P(si)$ indicates probability of i-th gray level in image.

This value will be equal to 8 in images which have been created completely random and will be considered as the ideal value. As much as the obtained results for entropy in a method is nearer to 8, it means that the prediction capacity of the method is less and the security rate of the method is higher. In the proposed algorithm, the obtained value is equal to 7.9911 that is very near to the ideal one, i.e. 8.This shows the high security of the method against attacks.

5 Decoding Method

One of the significant properties of the cryptography is the reversibility of encrypted image compared to the original one. With regard to the nature of the chaotic function, if there are key and initial value, we can regenerate a series of numbers that have been generated in the past. Therefore, we can regenerate the graph circle through using the numbers which have been regenerated by chaotic function. Then, we change the grey values of the pixels of encrypted image into the ones of the original image through using regenerated graph and numbers by employing XOR.

6 Conclusion

In this paper, a new method for encrypting image through using chaos function and graph theory has been proposed in order to complicate the encryption algorithm. As shown in the section allocated to the experimental results, this method showed robustness against all types of attacks such as decoding, statistical and brute-force attacks. The high value of entropy (7.9911) in this method shows the high performance of the proposed method.

References

1. Faridnia, S., enayatifar, R., sadeqi, H.: using chaotic function in image steganography. In: IEEE ICIME 2009, pp. 754–757 (2009)
2. Mitra, A., Subba Rao, Y.V., Prasanna, S.R.M.: A New Image Encryption Approach using Combinational Permutation Techniques. International Journal of Computer Science, pp: 1306–4428 (2006)
3. Chang, C.-C., Yu, T.-X.: Cryptanalysis of an encryption scheme for binary images. Pattern Recognition Letters, 1847–1852 (2002)
4. Joshi, M., Chandrashaker, K.S.: Color image encryption and decryption using fractional Fourier transform. Optics Communications, 811–819 (2007)
5. Roterman, Y., Porat, M.: Color image coding using regional correlation of primary colors. Image and Vision Computing, 637–651 (2007)
6. Alsultanny, Y.A.: Zandom-bit sequence generation from image data. Image and Vision Computing, 1178–1189 (2007)
7. Yen, J.-C., Guo, J.-I.: A New Chaotic Key-Based Design for Image Encryption and Decryption. In: Proceedings IEEE International Conference on Circuits and Systems, vol. 4, pp. 49–52 (2000)
8. Li, S., Zheng, X.: Cryptanalysis of a Chaotic Image Encryption Method. In: Proceedings IEEE International Symposium on Circuits and Systems, Scottsdale, AZ, USA, vol. 2, pp. 708–711 (2002)
9. Kwok, H.S., Tang, W.K.S.: A fast image encryption system based on chaotic maps with finite precision representation. Chaos, Solitons and Fractals, 1518–1529 (2007)
10. Behnia, S., Akhshani, A., Ahadpour, S., Mahmodi, H., Akhavan, A.: A fast chaotic encryption scheme based on piecewise nonlinear chaotic maps. Physics Letters A, 391–396 (2007)
11. Chen, G., Mao, Y.B., Chui, C.K.: A symmetric image encryption scheme based on 3D chaotic cat maps. Chaos, Solitons & Fractals, 74–82 (2004)
12. Mao, Y.B., Chen, G., Lian, S.G.: A novel fast image encryption scheme based on the 3D chaotic baker map. Int. Bifurcat Chaos, 544–560 (2004)
13. Pareek, N.K., Patidar, V., Sud, K.K.: Image encryption using chaotic logistic map. Image and Vision Computing, 926–934 (2006)
14. Shannon, C.E.: Bell Syst. Tech. J. 28, 656 (1949)

Robust Stamps Detection and Classification by Means of General Shape Analysis

Paweł Forczmański and Dariusz Frejlichowski

West Pomeranian University of Technology, Szczecin
Faculty of Computer Science and Information Technology
Żołnierska Str. 49, 71–210 Szczecin, Poland
{pforczmanski,dfrejlichowski}@wi.zut.edu.pl
http://en.wikipedia.org/wiki/West_Pomeranian_University_of_Technology

Abstract. The article presents current challenges in stamp detection problem. It is a crucial topic these days since more and more traditional paper documents are being scanned in order to be archived, sent through the net or just printed. Moreover, an electronic version of paper document stored on a hard drive can be taken as forensic evidence of possible crime. The main purpose of the method presented in the paper is to detect, localize and segment stamps (imprints) from the scanned document. The problem is not trivial since there is no such thing like "stamp standard". There are many variations in size, shape, complexity and ink color. It should be remembered that the scanned document may be degraded in quality and the stamp can be placed on relatively complicated background. The algorithm consists of several steps: color segmentation and pixel classification, regular shapes detection, candidates segmentation and verification. The paper includes also the initial results of selected experiments on real documents having different types of stamps.

1 Introduction

From a strictly technical point of view, rubber stamping, also called stamping, is a craft in which certain type of ink made of dye or pigment is applied to an image or pattern that has been carved, molded, laser engraved or vulcanized, onto a sheet of rubber. The ink coated rubberstamp is then pressed onto some type of medium such that the colored image has now been transferred to the medium. The medium is generally some type of fabric or paper. This kind of stamping has not changed for centuries (in fact it is as old as writing itself) and it is supposed that it will not change in the close future. Nowadays, when computer technology is present in various areas of life, the problem of computer crime is becoming more and more important. It covers both strictly electronic and traditional types of law-breakings. On the other hand, there are still many areas of life, where computers and digital media are employed only as tools and play just a supporting role. The most evident example of such domain is an area associated with official documents, identity cards, formal letters, certificates, etc. All these documents are being issued by formal authorities and are often

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 360–367, 2010.

in a form of a paper letter consisting of several typical elements: heading, body text, signatures and stamps which, from this historical point of view confirm its official character. In business environments, they are often used to provide supplemental information (date received/approved, etc). In other words, its main purpose is to authenticate a document which in many cases is a subject to forgery or tampering with help of modern computer means. In general, the process of forgery consists of the following steps: obtaining the original document, high resolution scanning, digital image manipulation and final printing. It is rather easy to recognize fake stamps, even if they are printed using ink-jet printers. This article addresses the problem, which is definitely not new, since the task of seal imprint identification on bank checks, envelopes, and transaction receipts have emerged from mid-1980s. On the other hand reliable recognition of stamps in the documents is not trivial and has not been solved till today [1,2,3]. The most advanced method found in the scientific literature is described in [2], where the authors present a stamp detection approach, which treats stamps as regions with analytically shaped contours, however this regions are limited to oval shapes only. The general motivation of the research presented in this paper is a need of semi-automatic computer software that is able to analyze an image and detect and localize different types of stamps in it. The application area of this kind of a system is broad, ranging form law-enforcement forces, law offices, official archives and any other institutions that utilize stamp.

2 Stamp Detection and Classification

2.1 Stamps Characteristics

All stamps placed on paper documents have specific characteristics which are derived from the process of stamping, These characteristics (shape, complexity, typical patterns) evolved into de-facto standards. The analysis of the problem shows that there are two main groups of stamps having its distinguishable properties: official stamps met mostly on official documents, unofficial stamps used as decoration. The first group (see Fig. 1) consists of regularly-shaped objects (ovals, squares, rectangles) with clearly visible text and mere ornaments. They are often colored red or blue and do not cover large areas. On the other hand, stamps from the second group (see Fig. 2) are more fancy, irregularly-shaped, with decorative fonts and complex patterns. It is a fundamental issue to define the features that can be employed to distinguish stamps from not-stamps and further between official and unofficial stamps. In this paper we focus on official stamps as it plays a meaningful role in practical tasks. The features which are used to describe stamps can be divided into two classes: spatial characteristics [4,5,6,7,8], including dimensions (proportions of dimensions), edge distributions, mean and variance of gradients, moment representation; color characteristics [9,10], which include color distribution is HSV and YC_bC_r color spaces. Beside these features it is always profitable to use stamp templates (as simplified images) to verify the detection and recognition stage. It is worth noticing that we

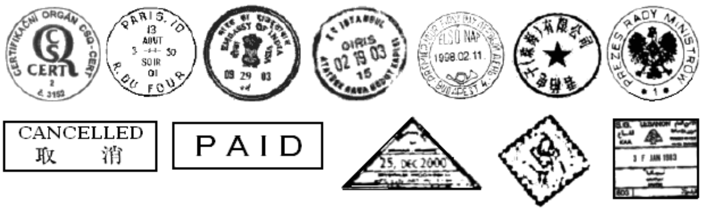


Fig. 1. Sample official stamps, often regular and without decorations

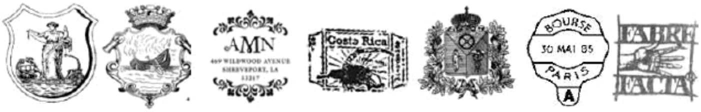


Fig. 2. Sample unofficial stamps, more complex, with many decorative motives

do not employ Hough transform to detect circles, since we deal also with rectangular, triangular and other stamps. Hence, the approach is much more flexible than one presented in [2].

2.2 Algorithm Overview

The algorithm of stamps processing is divided into several stages: detection, verification and coarse classification, which is depicted in Fig. 3. Detailed descriptions of each stage are presented in the following sections. The stamp detection uses color space transformation and projection method, stamp verification uses simple geometrical characteristics like object size and proportions of its dimensions while classification utilizes a novel approach to general shape analysis.

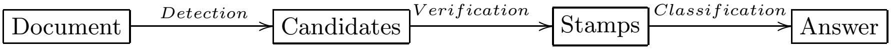


Fig. 3. Process of stamps retrieval from scanned document

2.3 Stamp Detection and Verification

An input image of a document has to be stored in a file with possibly lossless compression, high spatial resolution and full color range (24-bit RGB). First it is down-scaled to obtain low resolution representation (256×256 pixels) used for preliminary detection. The RGB image is then converted into YC_bC_r color space (ITU-R BT.709 standard). It is worth noticing that most popular stamps are often blue or red colored, thus we select C_b and C_r components only for further analysis (see Fig. 4 and Fig. 5). In each case presented below (Chinese taxation registration and Czech trade license), several potential areas are detected and passed to the verification/classification stage.

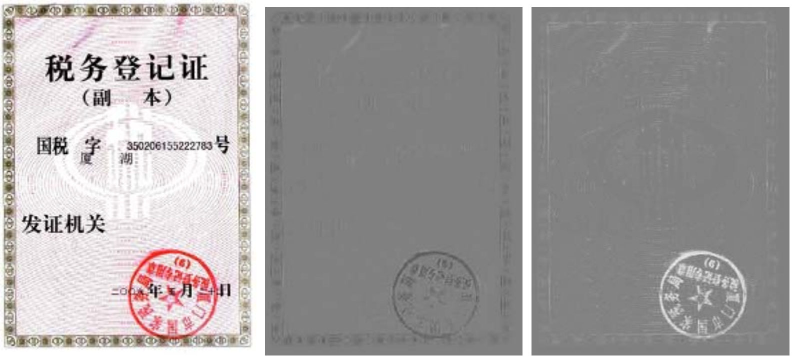


Fig. 4. Sample document containing a stamp (left), C_b and C_r components of its YC_bC_r representation

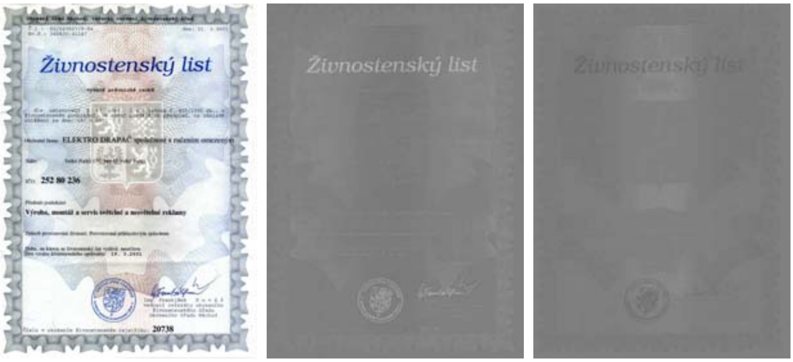


Fig. 5. Sample document containing a stamp (left), C_b and C_r component of its YC_bC_r representation (right)

For each matrix which represents C_b an C_r channel we perform projections in horizontal and vertical directions in order to find areas of high intensity. Sample projections for one of the test images are presented below, in Fig. 6. As it can be seen, areas occupied by possible stamps are represented by values higher than mean value (assumed to be a background value).

Next, the candidates for stamps are segmented and passed to the stage where the quasi-geometrical features are calculated. There are two general features used: object size and width to height proportion. The analysis of stamps collected for the experimental purposes, showed that object size should occupy not less than 5% and not more than 15% of the total image area. On the other hand, the proportion of width to height should be not less than 1/3 and not more that 3. This prevents the situation where relatively narrow objects are accepted. The algorithm is capable of detecting more than one stamp in a document as long as they occupy disjoint areas in the image. In case when the centers of two or more stamps are positioned close to each other, they are recognized as one stamp. In

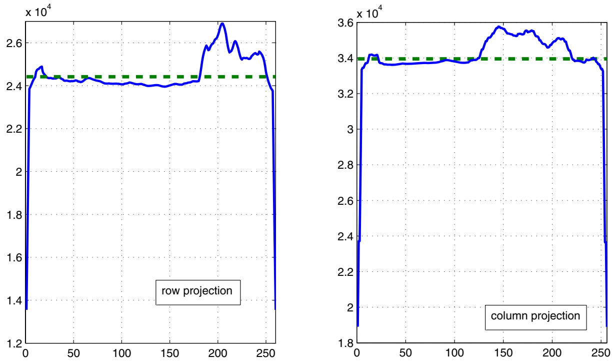


Fig. 6. Row projection (left) and column projection (right) and mean values (dashed line) of C_r component of the document image shown in Fig. 4

case when no stamp is detected (no areas of adequate color can be found), the algorithm ends. Finally, each extracted stamp is passed to the shape analysis stage in order to find its characteristic features.

2.4 Coarse Classification by Means of the General Shape Analysis

The general shape analysis (*GSA*) is a problem similar to shape classification or retrieval. There are two important differences. First of all, the number of model classes in the GSA is very small. This limited database covers only the most general objects — triangle, square, ellipse, etc. Secondly, the process of identification is different than in the classical recognition. In GSA the explored object does not have to belong to one of the base classes. In fact usually it does not. Thanks to this and the specifically selected instances in the database we can indicate the general information about a shape, e.g. how triangular, square, elliptical, etc. it is. The general shape analysis approach can be applied in all problems, where the initial coarse classification is desirable. This can be useful for example in retrieval of shapes, where firstly we are trying to identify a general class, and later work within this class only. Moreover, this process can be reiterated and at the consecutive stages the general shapes used may be more precise. However, not only retrieval is an example of the possible usage of GSA approach. It can be helpful also in problems, where we work using less precise description for processed objects. An example in this group is the usage of imprecise voice commands (e.g. 'find round blue objects') in any system of human-computer interaction. The mentioned properties of GSA allow us to use this approach in the second stage of the developed algorithm. Here we are trying to identify the general shape of an extracted stamp. The result of this step can help in indication of the character of a processed stamp. Basing on information about typical kinds of stamps we can pre-classify a particular one in order to decide if it has to be recognized more precisely. For example the much attention has to be put on round objects than for triangular ones, because usually the governmental, official stamps are round. Various shape description methods can be used in the GSA approach. However, depending on the particular application their different properties can be desirable. Usually, the property of generalization is useful, yet in the problem described in the paper, the ability of distinguishing the round, triangular and rectangular objects is especially helpful. That is why, amongst several algorithms explored in the problem of GSA (e.g. [11]), the most convenient ones are the polar methods. Amongst them the best results in the GSA problem was so far achieved by *Point Distance Histogram* (*PDH*). This conclusion was confirmed thanks to the comparison with human benchmark, obtained after analysis of inquiry forms, filled in by almost two hundred persons of different age and gender [11]. The *PDH* combines the advantages of histogram with the transformation of contour points into polar coordinates. Firstly the mentioned coordinates are derived (with $O = (O_x, O_y)$ as the origin of the transform) and put into two vectors Θ^i for angles and P^i for radii [11]:

$$\rho_i = \sqrt{(x_i - O_x)^2 + (y_i - O_y)^2}, \qquad \theta_i = atan\left(\frac{y_i - O_y}{x_i - O_x}\right). \tag{1}$$

The resultant values are converted into nearest integers [11]:

$$\theta_i = \begin{cases} \lfloor \theta_i \rfloor, & if\ \theta_i - \lfloor \theta_i \rfloor < 0.5 \\ \lceil \theta_i \rceil, & if\ \theta_i - \lfloor \theta_i \rfloor \geq 0.5 \end{cases}. \tag{2}$$

The next step is the rearrangement of the elements in Θ^i and P^i according to increasing values in Θ^i. This way we achieve the vectors Θ^j, P^j. For equal elements in Θ^j only the one with the highest corresponding value P^j is selected. That gives a vector with at most 360 elements, one for each integer angle. For further work only the vector of radii is taken — P^k, where $k = 1, 2, ..., m$ and m is the number of elements in P^k ($m \leq 360$). Now, the normalization of elements in vector P^k is performed [11]:

$$M = \max_k \{\rho_k\}, \qquad \rho_k = \frac{\rho_k}{M}, \tag{3}$$

The elements in P^k are assigned to r bins in histogram (ρ_k to l_k,[11]):

$$l_k = \begin{cases} r, & if\ \rho_k = 1 \\ \lfloor r\rho_k \rfloor, & if\ \rho_k \neq 1 \end{cases}. \tag{4}$$

In case of the above algorithm the L_2 norm can be used in order to derive the dissimilarity measure between two received shape descriptions h_1 and h_2 [12]:

$$L_2(h_1, h_2) = \sqrt{\sum_i ((h_1(i) - h_2(i))^2}. \tag{5}$$

The parameter r, determining the number of bins in the histogram representing a shape, was experimentally established as 25. There were only five models in the database, including the most popular stamp shapes, i.e. circle, ellipse, rectangle,

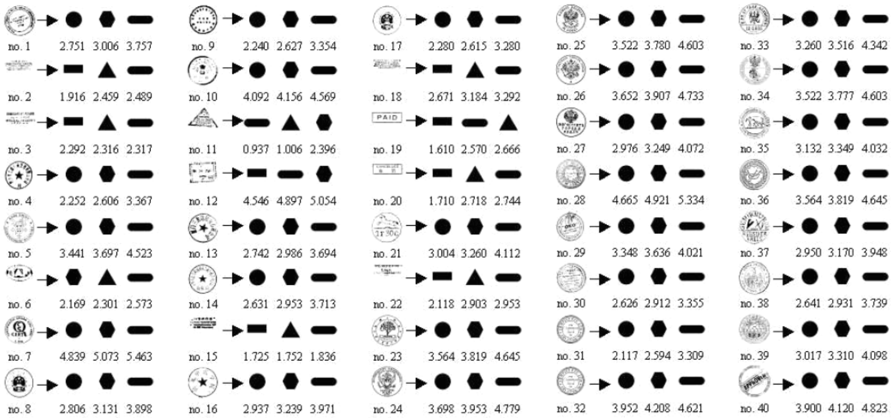


Fig. 7. The three general shapes the most similar to explored stamps, selected using *PDH* algorithm. The values of dissimilarity measure are also provided.

triangle and hexagon. Forty stamps of various size and rotation, extracted at the previous stages, were tested. The results are depicted in Fig. 7.

As usually in the case of general shape analysis [11] three base elements that are closest to the processed one are provided. As we can clearly see the *PDH* algorithm gives almost perfect answer. Only in two cases the results are wrong. The test object no. 6 is very small and indistinct and that could influence the results. In the second case only the first indication is incorrect. For test object no. 11 firstly the ellipse was pointed out. The proper answer — triangle — appeared in a second place. However, the difference between the similarity measures for the two first indicated general shapes were significantly small this time. During our experiments we discovered that the results of classification using *PDH* is actually not influenced by inaccurate segmentation, which shows as irregularities in shape contour. However, in very rare cases it might be corrected by blurring of input image after color separation.

3 Summary

In the paper we presented an approach which leads to the detection and coarse classification of stamps in the scanned documents. This problem has not been broadly discussed and analyzed so far. In fact, only one method, according to our best knowledge, can be found in the scientific literature [2]. Our algorithm is more complex and is composed of several steps, which include color conversion and separation, vertical and horizontal projection of pixel intensities and shape analysis and is more flexible, since detects not only oval stamps. The last stage of our approach was supported by the general shape analysis performed using Point Distance Histogram. That was helpful in general classification of stamps extracted from scanned documents. The selection of the method for this stage was intentional. First of all, we were looking for a method that will work better in discrimination between round, rectangular and triangular shapes versus another ones. It is important in our main problem of determining the potential criminal activity. In order to achieve this goal, the polar shape descriptors were suggested. They are by definition robust to many problems appearing when working with shapes. Basing on the experiments described in [11] *PDH* was pointed as the most effective algorithm within this groups in the problem of coarse classification. Its performance in this task was significantly better than in case of other polar shape descriptors. That was plainly confirmed by results presented and discussed in this paper. Our experiments conducted on 170 different scanned color documents shown that in more than 82% cases the stamp was successfully segmented and then in 93% successfully identified by means of the general shape analysis. Most of the errors were caused by poor quality of scanned documents as well as by a color different from the fixed standard. It might be easily improved by using more detailed color separation involving custom color models involving Gram-Schmidt orthonormalization procedure. The future research will be also focused on much more detailed stamp recognition, namely detection and recognition of texts present on stamps, which can have a very wide area of application.

References

1. Ueda, K., Nakamura, Y.: Automatic verification of seal impression patterns. In: Proc. 7th. Int. Conf. on Pattern Recognition, pp. 1019–1021 (1984)
2. Zhu, G., Jaeger, S., Doermann, D.: A robust stamp detection framework on degraded documents. In: Proceedings - SPIE The International Society For Optical Engineering, vol. 6067 (January 2006)
3. Zhu, G., Doermann, D.: Automatic document logo detection. In: The 9th International Conference on Document Analysis and Recognition (ICDAR 2007), pp. 864–868 (2007)
4. Pham, T.D.: Unconstrained logo detection in document images. Pattern Recognition 36, 3023–3025 (2003)
5. Zhang, D., Lu, G.: Review of shape representation and description techniques. Pattern Recognition 37, 1–19 (2004)
6. Loncaric, S.: A survey on shape analysis techniques. Pattern Recognition 31, 983–1001 (1998)
7. Mehtre, B.M., Kankanhalli, M.S., Lee, W.F.: Shape measures for content based image retrieval: a comparison. Information Proc. & Management 33, 319–337 (1997)
8. Wood, J.: Invariant pattern recognition: review. Pattern Recognition 29, 1–17 (1996)
9. Deng, Y., Manjunath, B.S., Kenney, C., Moore, M.S., Shin, H.: An efficient color representation for image retrieval. IEEE Transactions on Image Processing 10(1), 140–147 (2001)
10. Manjunath, B.S., Ohm, J.-R., Vasudevan, V.V., Yamada, A.: Color and texture descriptors. IEEE Transactions on Circuits and Systems for Video Technology 11, 703–715 (2001)
11. Frejlichowski, D.: An experimental comparison of seven shape descriptors in the general shape analysis problem. In: Campilho, A., Kamel, M. (eds.) ICIAR 2010. LNCS, vol. 6111, pp. 294–305. Springer, Heidelberg (2010)
12. Miklasz, M., Aleksiun, P., Rytwinski, T., Sinkiewicz, P.: Image recognition using the histogram analyser. Multimedia and Intelligent Techniques 1, 74–86 (2005)

Pre-processing, Extraction and Recognition of Binary Erythrocyte Shapes for Computer-Assisted Diagnosis Based on MGG Images

Dariusz Frejlichowski

West Pomeranian University of Technology, Szczecin
Faculty of Computer Science and Information Technology
Żołnierska 49, 71-210, Szczecin, Poland
dfrejlichowski@wi.zut.edu.pl

Abstract. The paper presents an approach for computer-assisted diagnosis of some diseases (e.g. anaemia, malaria), which are caused by deformations of erythrocyte shapes. In the method firstly the thresholding of the input MGG image is performed, using modified thresholding based on fuzzy measures. Secondly, every cell is localised and extracted. Only the red blood cells are taken for later processing. Using the template matching approach and shape description algorithm every extracted erythrocyte is assigned to one of the twelve classes. Basing on the knowledge about the number of particular unaffected and affected red blood cells the diagnosis can be made. Hence, the possibilities of automatic diagnosis are discussed to stress the potential application of the method.

1 Introduction

Diagnosis of some human diseases can be based on morphotic analysis of the blood. When working with the digital version of microscopic image various features can be used. From the image description point of view the most suitable are colour, size, texture and shape. Amongst them shape brings especially important information, because it is visibly different for healthy and affected erythrocytes. Moreover, usually modification of this cell feature leads to a disease (e.g. malaria and anaemia). It comes from the fact that deformed erythrocytes cannot deliver oxygen properly and blood circulation is non-regulated.

The algorithms presented in this paper were explored using digital microscopic images of human blood stained using *MGG (May-Grunwald-Giemsa)* method. Because purple and pink are dominating colours in such images firstly they were converted to grayscale. In result the background is very light in comparison to the important objects within (e.g. cells). The most problematic are blurred objects, mainly on the edges and occluded cells.

The task of automatic diagnosis basing on erythrocyte shape is not new. However it is not as popular as other medical applications of image analysis and recognition. An example of the approach for identification of a particular

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 368–375, 2010.

disease was presented in [1], where complete scheme was proposed, based on some histogram features. However, usually various particles are analysed. For example in [2] 12 categories of particles in human urine were classified. As for the blood analysis, leukocytes identification is more popular (e.g. [3,4]), however research interest on automatic recognition of red blood cells is recently arising. Apart from the mentioned paper ([1]) three another research examples are worth mentioning. In [5] deformable templates were used. In [6] morphological operators were applied. Some preliminary results on application of polar-logarithmic transform to the discussed problem were provided in [7].

The presented in the paper method for computer-assisted diagnosis basing on the MGG images can be divided into three main stages. The first one is the binarisation of the input image performed using modified thresholding based on fuzzy measures. The second is the extraction and selection of a single red blood cell, and the last one is the shape recognition of each extracted erythrocyte contour.

2 Preprocessing

Using the grayscale *MGG* images, the modified histogram thresholding is performed to achieve a binary image for the further work. This process starts with the construction of the histogram, represented as a function $h(l_k)$:

$$h(l_k) = \sum_{k=1}^{m} b(k, l_k), \tag{1}$$

where:

$$b(k, l_k) = \begin{cases} 1, & \text{if } k = l_k \\ 0, & \text{if } k \neq l_k \end{cases} . \tag{2}$$

Afterwards, the achieved representation is smoothed. Originally ([8]) this process was performed through bins averaging using ‘large window’. However, the authors did not specify the minimal sufficient length of the window. That is why here this length was derived using the experimentally established formula:

$$c(j) = \frac{\sum\limits_{i=j-m}^{j+m} h(i)}{2m+1}, \tag{3}$$

where:
j — number of a bin,
$c(j)$ — averaged histogram value for bin,
$h(i)$ — histogram value before averaging,
m — number of bins taken in left and right neighborhood of j-th bin.

The calculation of the threshold value is simple and fast. It comes from the properties of the histogram achieved for *MGG* in grayscale. It is composed of two distinct peaks. The first one covers the cytoplasm, and the second — light background pixels. The appropriate threshold should be established between them ([9]).

The idea is to use the neighbourhood of the maximal bin in histogram. Basing on the performed experiments the number of distinct gray-levels in a particular image has to be taken into consideration. If it is higher than 150, the threshold (t) is derived as the minimal value in the interval:

$$t \in (c_{\max} - \left\lfloor \frac{v}{4} \right\rfloor, c_{\max}), \tag{4}$$

where:
c_{max} — number of the highest bin,
v — number of gray-levels found in particular image (number of non-zero bins in histogram).

If the number of non-zero bins in histogram is smaller than 150, the parameter t is calculated as the minimal value in other interval:

$$t \in (c_{\max} - \left\lfloor \frac{v}{4} \right\rfloor, c_{\max} - 20). \tag{5}$$

An exemplary result of the above procedure is depicted in fig. 1. As we can see, the black pixels are achieved for background, and the white ones for marked cells.

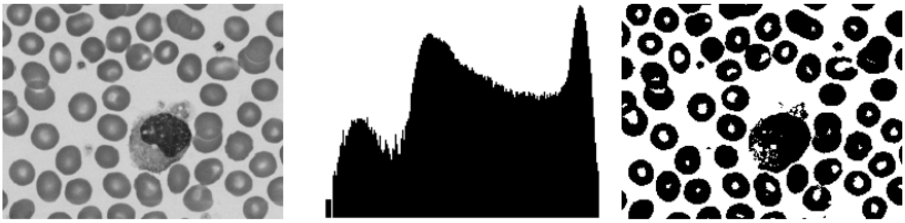


Fig. 1. An example of the threshloding result achieved using the method described in the paper — from left to right: the MGG image, the histogram achieved for it and the thresholded image

In the next stage, only the objects entirely placed in images are considered. Therefore, the ones placed on edges are removed and later all remaining cells (black shapes on white background) are localised by tracing the outline of each separate object in image. Afterwards, leukocyte and thrombocyte shapes are removed, basing on the knowledge about the typical sizes of blood particles: leukocytes are significantly larger and thrombocytes smaller. This is usually, but not always, true. Hence, the information about grayscale histogram for particular cell is used, because red blood cell is lighter than the other two. After the rejection of leukocytes and thrombocytes, the extraction of each object starts. Simply, the first foreground pixel and the maximal boundaries of the object to which it belongs are found. Finally, the extracted pixels are replaced with background pixels. Thanks to this they are not considered later. The process ends, when there is no foreground pixels.

The above stages gave the coordinates for every erythrocyte in the binary image. For later recognition, the outline of an object is used. However, the existence of lighter part within the cell is also considered. If it exists, it is extracted and used later as well.

Basing on the template matching approach, the extracted contour shape is represented using a shape descriptor and matched with the base elements (represented in the same manner) in order to indicate the most similar one. The highest value of similarity measure indicates the recognition result.

3 Shape Representation and Matching

Three shape description methods were explored. Because the erythrocyte's shape is usually round, the polar methods were chosen.

The first approach, *UNL-Fourier (UNL-F*, [10]), is composed of two transformations. Firstly, the transformation of boundary points to polar coordinates (more precisely, *UNL-transform*) is performed and secondly *2D Fourier transform*. The resultant shape description is invariant to scaling, translation and rotation. The representation is also robust to some degree of noise.

The first stage is the *UNL-transform* ([10]). It uses complex representation of Cartesian coordinates for points and parametric curves in discrete manner:

$$z(t) = (x_1 + t(x_2 - x_1)) + j(y_1 + t(y_2 - y_1)), \qquad t \in (0,1), \tag{6}$$

where $z_1 = x_1 + jy_1$ and $z_2 = x_2 + jy_2$ (complex numbers) and z_i means point with coordinates x_i, y_i.

The centroid O is firstly calculated:

$$O = (O_x, O_y) = (\frac{1}{n}\sum_{i=1}^{n} x_i, \frac{1}{n}\sum_{i=1}^{n} y_i), \tag{7}$$

and the maximal Euclidean distance between points and centroid is found:

$$M = \max_i\{\|z_i(t) - O\|\}, \qquad \forall i = 1...n, \qquad t \in (0,1). \tag{8}$$

Now coordinates are transformed:

$$U(z(t)) = R(t) + j \times \theta(t) = \frac{\|z(t) - O\|}{M} + j \times atan(\frac{y(t) - O_y}{x(t) - O_x}). \tag{9}$$

The discrete version is formulated as follows:

$$\begin{aligned} U(z(t)) = {} & \tfrac{\|(x_1+t(x_2-x_1)-O_x)+j(y_1+t(y_2-y_1)-O_y)\|}{M} \\ & +j \times atan(\tfrac{y_1+t(y_2-y_1)-O_y}{x_1+t(x_2-x_1)-O_x}). \end{aligned} \tag{10}$$

The parameter i is discretized in the interval [0,1] with significantly small steps. Derived coordinates are put into matrix, in which row corresponds to the distance from centroid, and column — to the angle. The obtained matrix is 128 × 128 pixels size. In fact, that gives another binary image. Thanks to this *2D Fourier transform* can be applied to it, as in the equation below. Usually the absolute spectrum of the *Fourier description* is used ([11]):

$$C(k,l) = \frac{1}{HW} \left| \sum_{h=1}^{H} \sum_{w=1}^{W} P(h,w) \cdot e^{(-i\frac{2\pi}{H}(k-1)(h-1))} \cdot e^{(-i\frac{2\pi}{W}(l-1)(w-1))} \right|, \quad (11)$$

where:
H, W — height and width of the image in pixels,
k — sampling rate in vertical direction ($k \geq 1$ and $k \leq H$),
l — sampling rate in horizontal direction ($l \geq 1$ and $l \leq W$),
$C(k,l)$ — value of the coefficient of discrete Fourier transform in the coefficient matrix in k row and l column,
$P(h,w)$ — value in the image plane with coordinates h, w.

The *Log-Pol-F* ([7]) was the second approach examined. The *polar-logarithmic transform* is used here (with the centroid as an origin), and, afterwards, the *Fourier transform.* Thanks to the similarity to the previous descriptor the *Log-Pol-F* has similar properties.

As in the earlier approach, the method starts with the calculation of the centroid O (see eq.7). Then, the transform can be formulated as follows:

$$\rho_i = \log(\sqrt{(x_i - O_x)^2 + (y_i - O_y)^2}), \qquad 0 \leq \rho_i \leq \rho_{\max}, \quad (12)$$

$$\varpi_i = arctan(\frac{y_i - O_y}{x_i - O_x}), \qquad 0 \leq \varpi_i \leq 2\pi. \quad (13)$$

The first equation allows to receive the *logarithmic-radius coordinates*, and the second the *angular* ones. As in the previous shape descriptor the last stage is the calculation of the *Fourier transform* (see eq.11) and selection of the most important coefficients.

The *Point Distance Histogram* ([12]) was the last algorithm explored. The descriptor is based on a combination of *polar coordinates* and *histogram.*

The *PDH* starts with the derivation of *polar coordinates* and insertion of them into two vectors: Θ^i for angles and P^i for radii:

$$\rho_i = \sqrt{(x_i - O_x)^2 + (y_i - O_y)^2}, \qquad \theta_i = atan\left(\frac{y_i - O_y}{x_i - O_x}\right), \quad (14)$$

The obtained values are converted into nearest integers:

$$\theta_i = \begin{cases} \lfloor \theta_i \rfloor, & if\ \theta_i - \lfloor \theta_i \rfloor < 0.5 \\ \lceil \theta_i \rceil, & if\ \theta_i - \lfloor \theta_i \rfloor \geq 0.5 \end{cases} . \quad (15)$$

The elements in Θ^i and P^i are rearranged, according to increasing values in Θ^i, and denoted as Θ^j, P^j. If some elements in Θ^j are equal, only the one with the highest corresponding value P^j is selected. That gives a vector with at most 360 elements, one for each integer angle. For further work only the vector of radii is needed. We denote it as P^k, where $k = 1, 2, ..., m$ and m is the number of elements in P^k ($m \leq 360$). Now, the normalization of elements in vector P^k is performed:

$$M = \max_k \{\rho_k\}, \qquad \rho_k = \frac{\rho_k}{M}, \tag{16}$$

The elements in P^k are assigned to bins in histogram (ρ_k to l_k):

$$l_k = \begin{cases} r, & if\ \rho_k = 1 \\ \lfloor r\rho_k \rfloor, & if\ \rho_k \neq 1 \end{cases}, \tag{17}$$

The parameter r denotes the predetermined number of bins and it can vary, depending on the application. In our case it was experimentally established as equal to 50. The values in bins are now normalized according to the highest one. This is the final step. The achieved histogram is the representation of a shape. It is invariant to translation, scaling and rotation. It also achieved good results in indication of small differences between similar objects.

For every explored experimentally method of erythrocyte shape description the template matching approach for classification was used. Obviously, firstly the five descriptions for twelve classes of RBCs were derived and stored as the templates. The particular object under identification was also represented using a shape description method and the similarity measure between it and templates indicated the recognized variant of RBC. For this purpose the $C1$ metric was used ([13]):

$$C1(X,Y) = 1 - \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} |x_{ij} - y_{ij}|}{\sum_{i=1}^{M}\sum_{j=1}^{N} (|x_{ij}| - |y_{ij}|)}, \tag{18}$$

where:
X, Y — descriptions matched,
M, N — height and width of the descriptions.

The experiments were performed for 55 digital microscopic images, magnified 1000 times. The detailed results for each erythrocyte class separately are provided in Table 1. As we can see, the *UNL-F* shape descriptor gave the best result (93% recognition rate). *Log-Pol-F* was slightly worse (91% RR), and *PDH* achieved the recognition rate equal to 84%. Basing on those results we can conclude that robustness to noise (both descriptors applying the Fourier transform) is more useful in our problem than the emphasis of small differences between shapes (*PDH*).

Additionally, considering future application in automatic diagnosis, the confidence value was experimentally established. In case of *UNL-F* (the best method amongst the explored) if the similarity measure (achieved using metric *C1*) was

Table 1. Average recognition rates (%) achieved for each class of erythrocytes

Class	UNL-F	Log-Pol-F	PDH
1. schistocyte	98	96	92
2. dacrocyte	100	96	90
3. acantocyte	90	92	82
4. echinocyte	84	86	76
5. ovalocyte	94	90	82
6. normocyte	90	88	80
7. stomatocyte	96	92	82
8. Mexican hat cell	94	94	86
9. spherocyte	90	86	82
10. leptocyte	92	84	82
11. annular erythrocyte	90	90	84
12. drepanocyte	100	98	88
TOTAL	93	91	84

lower than 0.71, the resultant cell was denoted as ‘doubtful’. It means that it cannot be assigned properly to any base class and is not taken into consideration.

4 Conclusions, Discussion and Future Plans

Considering results described in the previous section the *UNL-F* descriptor was chosen for representation of erythrocyte shapes and *C1* metric for matching. As it was stated, the recognition rate was not perfect, however it was enough to perform some initial experimental diagnosis.

The large number of objects in the image is a very useful property here. It simplifies the definition of the decision rules. The recognition rate has not to be perfect, what is almost impossible for real images. The experiments performed in ([12]) have shown that even the results above 80 percent are sufficient to make a correct diagnosis. This comes from the mentioned number of cells under analysis. Because for *MGG* images we are identifying several hundreds of them, the indication of particular abnormal erythrocyte shapes is easy.

The discussed issue was verified using 55 *MGG* images diagnosed by human and by a prototype in *Matlab*, using the presented approaches. In all cases the diagnoses results were the same. It confirmed that even if the recognition rates for single shapes are not ideal, the large number of them in the image gives correct results as a whole. Moreover, affected *RBCs* are easier to find, thanks to their visible change in shape in comparison with the normal ones.

The planned works cover three main aspects. Above all, the automatic diagnosis based on the approach presented in this paper has to be tested on much bigger collection of images. Secondly, the precise formulation of all decision rules for the diagnosis will be stated. Finaly, in order to increase the number of diseases diagnosed, not only erythrocytes, but also leukocytes and thrombocytes can be considered.

Acknowledgments. The author of this paper wishes to thank gratefully MSc J. Mlynarczykowska for significant help in exploring the proposed methods.

References

1. Díaz, G., González, F.A., Romero, E.: A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic imageses. Journal of Biomedical Informatics 42(2), 296–307 (2009)
2. Ranzato, M., Taylor, P.E., House, J.M., Flagan, R.C., LeCun, Y., Perona, P.: Automatic recognition of biological particles in microscopic images. Pattern Recognition Letters 28(1), 31–39 (2007)
3. Song, X.B., Abu-Mostafa, Y., Sill, J., Kasdan, H., Pavel, M.: Robust image recognition by fusion of contextual information. Information Fusion 3(4), 277–287 (2002)
4. Sabino, D.M.U., da Fontoura Costa, L., Rizzatti, E.G., Zago, A.M.: A texture approach to leukocyte recognition. Real-Time Imaging 10(4), 205–216 (2004)
5. Bronkorsta, P.J.H., Reinders, M.J.T., Hendriks, E.A., Grimbergen, J., Heethaar, R.M., Brakenhoff, G.J.: On-line detection of red blood cell shape using deformable templates. Pattern Recognition Letters 21(5), 413–424 (2000)
6. Di Ruberto, C., Dempster, A., Khan, S., Jarra, B.: Analysis of infected blood cell images using morphological operators. Image and Vision Computing 20(2), 133–146 (2002)
7. Luengo-Oroz, M.A., Angulo, J., Flandrin, G., Klossa, J.: Mathematical Morphology in Polar-Logarithmic Coordinates. In: Marques, J.S., Pérez de la Blanca, N., Pina, P. (eds.) IbPRIA 2005. LNCS, vol. 3523, pp. 199–205. Springer, Heidelberg (2005)
8. Sanei, S., Lee, T.K.M.: Cell Recognition based on PCA and Bayesian classification. In: Proc. ICA 2003, Nara, Japan, pp. 239–243 (2003)
9. Frejlichowski, D.: Binarisation of MGG Images Using Modified Histogram Thresholding. Polish Journal of Environmental Studies 17(3b), 108–112 (2008)
10. Rauber, T.W., Steiger-Garcao, A.S.: 2-D form descriptors based on a normalized parametric polar transform (UNL transform). In: Proc. MVA 1992 IAPR Workshop on Machine Vision Applications (1992)
11. Kukharev, G.: Digital Image Processing and Analysis (in Polish). Szczecin University of Technology Press (1998)
12. Frejlichowski, D.: The Point Distance Histogram for Analysis of Erythrocyte Shapes. Polish Journal of Environmental Studies 16(5b), 261–264 (2007)
13. Lam, K.M., Yan, H.: An Analytic-to-Holistic Approach for Face Recognition Based on a Single Frontal View. IEEE Trans. on PAMI 20(7), 673–686 (1998)

Analysis of Four Polar Shape Descriptors Properties in an Exemplary Application

Dariusz Frejlichowski

West Pomeranian University of Technology, Szczecin
Faculty of Computer Science and Information Technology
Żołnierska 49, 71-210, Szczecin, Poland
dfrejlichowski@wi.zut.edu.pl

Abstract. The paper presents the results of a comparison of four shape representation techniques that use the transformation of the points from Cartesian to polar coordinates. An exemplary application was selected — the recognition of characters extracted from polish car license plates. The main goal was the selection of the best polar shape descriptor basing on the experiments performed as well as the discussion on the application of the algorithms based on polar transform in the mentioned problem. The results of the experiment presented in the paper indicate that amongst used methods the Point Distance Histogram (PDH) works the best. The results achieved using the UNL and UNL-Fourier transforms are worse, yet still acceptable.

1 Introduction

The problem of automatic license plates recognition (*ALPR, LPR*) can be easily found in many crucial applications nowadays, e.g. traffic and parking management, tolling, access control, speed limit enforcement, identification of stolen cars. Many commercial products have been proposed and successfully used.

A system for license plate recognition is usually composed basing on five main stages: image acquisition, pre-processing, segmentation of a license plate, character separation, and character recognition. The paper focuses on the last stage, when we are trying to identify the extracted characters, one by one. This can be considered as a classification with the number of classes equal to the number of various characters that can occur in a particular license plate. Because the problem is important and practically useful, there have been many approaches to it developed so far. That is why one can find many examples of methods applied to the *LPR* problem, e.g. neural networks ([1,2,3,4]), genetic algorithms ([5]), fuzzy logic ([4]), Support Vector Machines ([6]), Hidden Markov Model ([7]), synergetic approach ([8]), combination of sub-classifiers ([9]), histograms ([10]), a vertical edge matching based algorithm ([11]), matching using simple templates ([12]), Hausdorff distance ([13]).

The template matching is one of the general methods that can be used in the recognition of characters extracted from license plates. In this approach one instance for a class is present in the database. It is a template. Usually it is properly

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 376–383, 2010.

prepared, e.g. transformed using a shape description method. Obviously, during the process of matching the object under recognition has to be represented in the same way. Thanks to this the classification can be easily come down to the matching between the object being recognized and all templates. The one with the highest similarity measure (or smallest dissimilarity measure) is selected as the recognized one.

Amongst the several dozens of algorithms the very useful properties can be found in the polar shape descriptors. They apply the transformation from Cartesian into polar coordinates. They are invariant to translation (thanks to the calculation of new coordinates according to a particular point, usually inside the shape) and scaling (if the normalization is applied). The rotation becomes a circular shift after the transform, what can be easily solved using another step in the algorithm, e.g. Fourier transform or a histogram. The polar shape descriptors can be obtained quickly and easily. Having in mind all the presented advantages, in the experiments four algorithms based on the polar transform of the coordinates of a contour points were utilized: *UNL*, named after the *Universidade Nova de Lisboa* ([14]), *UNL-Fourier* (*UNL-F*, [15]), *Point Distance Histogram* (*PDH*, [16]) and *Log-Pol* (*logarithmic-polar transform*, [17]). They will be described briefly in the next section.

2 Description of the Polar Shape Descriptors Used in the Experiments

The *UNL* (*Universidade Nova de Lisboa*) descriptor is the first algorithm selected for the recognition of characters extracted from polish car license plates. It uses a transform with the same name ([14]). The approach starts with the complex representation of Cartesian coordinates for points and usage of parametric curves in discrete manner ([14]):

$$z(t) = (x_1 + t(x_2 - x_1)) + j(y_1 + t(y_2 - y_1)), \qquad t \in (0, 1), \tag{1}$$

where $z_1 = x_1 + jy_1$ and $z_2 = x_2 + jy_2$ are complex numbers and z_i is a point with coordinates x_i, y_i. The parameter n denotes the number of points in the contour.

The centroid O is an origin of the transform and it is calculated using the formula ([14]):

$$O = (O_x, O_y) = (\frac{1}{n}\sum_{i=1}^{n} x_i, \frac{1}{n}\sum_{i=1}^{n} y_i). \tag{2}$$

Now the maximal Euclidean distance between points and centroid is found ([14]):

$$M = \max_i\{\|z_i(t) - O\|\}, \qquad \forall i = 1...n, \qquad t \in (0, 1). \tag{3}$$

The coordinates are transformed ([14]):

$$U(z(t)) = R(t) + j \times \theta(t) = \frac{\|z(t) - O\|}{M} + j \times atan(\frac{y(t) - O_y}{x(t) - O_x}). \tag{4}$$

The discrete version of the above is formulated as follows ([14]):

$$\begin{aligned} U(z(t)) = & \tfrac{\|(x_1+t(x_2-x_1)-O_x)+j(y_1+t(y_2-y_1)-O_y)\|}{M} \\ & + j \times atan(\tfrac{y_1+t(y_2-y_1)-O_y}{x_1+t(x_2-x_1)-O_x}). \end{aligned} \tag{5}$$

The parameter i is discretized in the interval $[0,1]$ with significantly small steps ([15]). Derived coordinates are put into matrix, in which row corresponds to the distance from centroid, and column — to the angle. The obtained matrix is 128×128 pixels size.

Because after the *UNL-transform* we achieve 2-dimensional binary image again, the author of the approach proposed the usage of the 2D Fourier transform as a next step ([15]). That gave a very effective descriptor, called shortly the *UNL-F*, which is not only invariant to rotation, translation and scaling, but also robust to some level of noise. This method is the second descriptor used in the experiments. The required 2D Fourier transform can be calculated using the formula ([18]):

$$C(k,l) = \frac{1}{HW} \left| \sum_{h=1}^{H} \sum_{w=1}^{W} P(h,w) \cdot e^{(-j\frac{2\pi}{H}(k-1)(h-1))} \cdot e^{(-j\frac{2\pi}{W}(l-1)(w-1))} \right|, \tag{6}$$

where:
H, W — height and width of the image in pixels,
k — sampling rate in vertical direction ($k \geq 1$ and $k \leq H$),
l — sampling rate in horizontal direction ($l \geq 1$ and $l \leq W$),
$C(k,l)$ — value of the coefficient of discrete Fourier transform in the coefficient matrix in k row and l column,
$P(h,w)$ — value in the image plane with coordinates h, w.

The *Point Distance Histogram* ([16]) is the third algorithm explored in the problem. As in the previous methods it uses the centroid $O = (O_x, O_y)$ as the origin of the transform. The resultant coordinates are put into two vectors Θ^i for angles (in degrees) and P^i for radii ([16]):

$$\rho_i = \sqrt{(x_i - O_x)^2 + (y_i - O_y)^2}, \qquad \theta_i = atan\left(\frac{y_i - O_y}{x_i - O_x}\right). \tag{7}$$

The resultant values in θ_i are converted into nearest integers ([16]):

$$\theta_i = \begin{cases} \lfloor \theta_i \rfloor, & if\ \theta_i - \lfloor \theta_i \rfloor < 0.5 \\ \lceil \theta_i \rceil, & if\ \theta_i - \lfloor \theta_i \rfloor \geq 0.5 \end{cases}. \tag{8}$$

The next step is the rearrangement of the elements in Θ^i and P^i according to the increasing values in Θ^i. This way we achieve the vectors Θ^j, P^j. For equal elements in Θ^j only the one with the highest corresponding value P^j is selected.

That gives a vector with at most 360 elements, one for each integer angle. For further work only the vector of radii is taken — P^k, where $k = 1, 2, ..., m$ and m is the number of elements in P^k ($m \leq 360$). Now, the normalization of elements in vector P^k is performed ([16]):

$$M = \max_k \{\rho_k\}, \qquad \rho_k = \frac{\rho_k}{M}, \tag{9}$$

The elements in P^k are assigned to r bins in histogram (ρ_k to l_k,[16]):

$$l_k = \begin{cases} r, & if\ \rho_k = 1 \\ \lfloor r\rho_k \rfloor, & if\ \rho_k \neq 1 \end{cases}. \tag{10}$$

The *Log-Pol* descriptor is the last method used in the experiments. It is based on the combination of the logarithmic and polar transforms. Again it starts with the calculation of the centroid $O = (O_x, O_y)$. Using it as the origin of the transform the logarithmic-polar coordinates are derived ([17]):

$$\rho_i = log(\sqrt{(x_i - O_x)^2 + (y_i - O_y)^2}), \qquad \omega_i = atan\left(\frac{y_i - O_y}{x_i - O_x}\right). \tag{11}$$

After the transform we receive an image with the angular coordinates placed on the vertical axis and the logarithm of the radii coordinates on the horizontal one.

3 Discussion on the Behaviour of the Explored Algorithms

The polar shape description algorithms are applied to many problems of machine vision, when the shape of an object is used as a feature. In this paper only the methods using contour of an objects are explored, however there is also a small class of algorithms working on the shape as a whole, with its interior. One of the most popular methods belonging to this group is the *Generic Fourier Descriptor* (*GFD*, [19]).

The most important property of shape description techniques in the recognition of objects extracted from a digital image is the invariance to particular deformations of a shape. Amongst them the invariance to several affine transformations is absolutely fundamental. The robustness to some level of noise is also useful. Obviously, in other machine vision applications, some other properties can be desirable instead of the above. For example in specific shape retrieval problems one can establish very rigid limitations that can make the invariance to selected transforms less important. That suitable case is rather rare in real recognition problems.

One of the first steps in all polar shape description algorithms is the calculation of the origin of the transform. Usually the centroid performs this role. The calculation of new polar coordinates from the initial Cartesian ones according to the origin gives in result invariance to translation, the first affine transform.

The change of scale is the second problem. In *UNL*, *UNL-F* and *PDH* it is easily solved through the normalization of the distances from points to the origin according to the highest one. In the fourth algorithm, *Log-Pol*, the change of size in the original shape becomes horizontal shift in the transformed image, according to the autosimilarity property of the exponential function ([17]). The above and the *UNL* descriptor are not directly invariant to rotation. Rotation of the object in Cartesian coordinates results in vertical cyclic shift in the polar ones. This can be solved using additional step in an algorithm. In *UNL-F* the Fourier transform is used, and in *PDH* — the histogram. Moreover, the *UNL-F* descriptor is robust to some level of noise, thanks to the usage of the Fourier transform.

The basic usage of the polar transform in shape representation can give in result a description that can be limited and strongly dependent on the character of explored objects. For example, the so-called 'signatures' that simply contain the vector of distances from the origin to the particular points, ordered accordingly to the tracing the contour, are better suited for star-shaped objects. Moreover, they can work worse for similar objects, like the letters and digits extracted from the car license plates. This was especially visible in case of *Log-Pol*, because this approach is the most similar to the simple shape signatures. However, other methods selected for experiments are less sensitive to this problem. In case of *UNL* and *UNL-F* the specific way of composing the resultant description is helpful. Instead of ordering basing on the tracing the contour, the resultant matrix contains the distance from the origin ordered according to the increasing angular component. In *PDH* the usage of histogram is useful. Still, we have to consider carefully the number of bins selected for shape representation. If it is to small or to large, it will influence the results. Hence, here the experimentally established value ($r = 50$) was used. The lower ambiguity in character representation for *UNL* and *PDH* is pictorially presented in Fig. 1. In the same image we can notice the problem with the *Log-Pol* descriptor. The achieved representations of different characters are more similar to each other than in the previous two algorithms.

The described algorithms were tested on shapes represented as contours, localised and extracted from real images, obtained in various weather and imaging conditions (e.g. snow, rain, blurred images). The problem of car and license plate localisation was not considered. The part of an image with single digit or letter was treated as input. Only the outer contour was used. The number of test instances was equal to 280, while the number of templates was equal to 35, what is the number of characters used in polish car license plates, according to the polish norm PN-S-73200. The process of character identification was based on the template matching approach. In this case the templates as well as the test objects are represented using the same shape description method and then a description of a test object is matched with all descriptions of base objects using a similarity measure.

The average recognition rates achieved using particular polar shape description methods were as follows: the *PDH* — 96%, the *UNL* — 89%, the *UNL-F* —

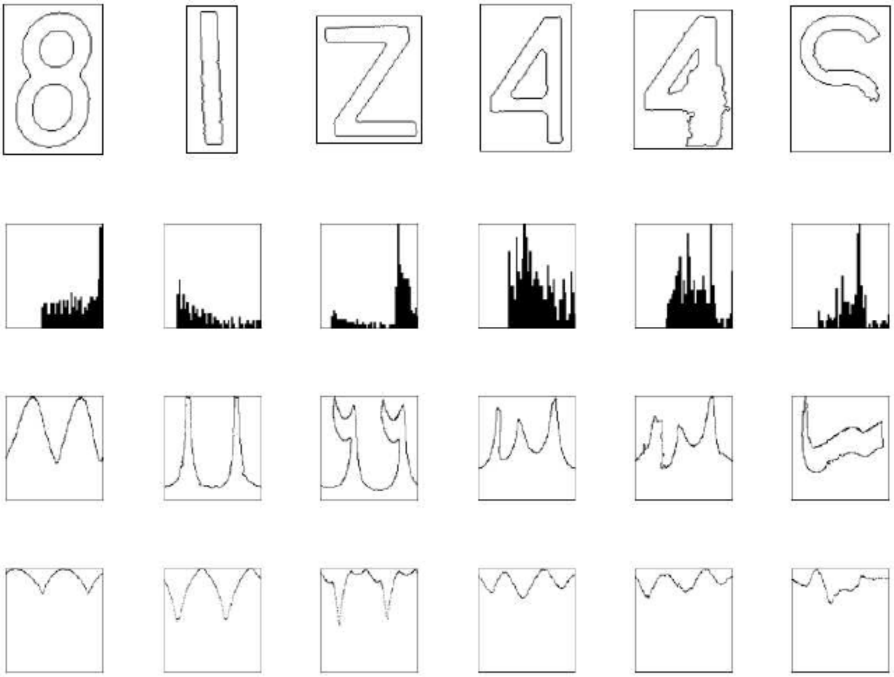


Fig. 1. The pictorial representations of exemplary shape descriptions achieved for explored characters (first row) using PDH (second row), UNL (third row) and Log-Pol (last row)

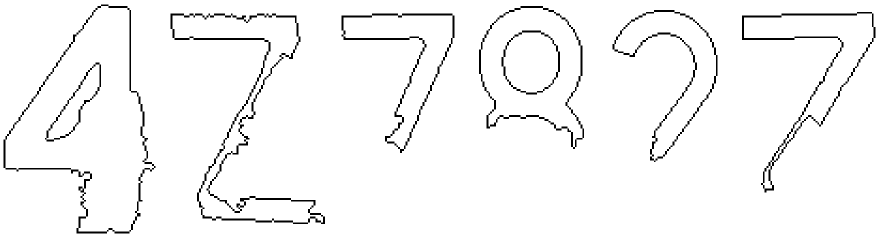


Fig. 2. Examples of strongly distorted characters used in the experiments

91% and the *Log-Pol* — 82%. As we can see the results of the *PDH* descriptor are very promising. The achievements of other three polar algorithms explored are worse. It was caused by some specific problems. An example is the similarity between some characters, e.g. the letter O and the digit 0. However, the most important were the noise on a contour and occlusions caused for example by weather conditions, dirt, and plate's devastation. Several examples of such problematic objects are provided in Fig. 2.

4 Conclusions

In the paper four polar shape description methods were analysed in an exemplary problem — recognition of characters extracted from polish car license plates. The template matching approach was used, and the description of a shape contour represented an object. Amongst the algorithms explored the best result during the experiments was achieved using the *PDH* shape descriptor. The results of application of the *UNL* and *UNL-F* were worse, yet still acceptable, considering the strong deformations of explored characters as well as some specific problems, e.g. the similarity between some of the letters and digits. The worse result was achieved using the *Log-Pol* algorithm.

References

1. Youssef, S.M., AbdelRahman, S.B.: A Smart Access Control Using an Efficient License Plate Location and Recognition Approach. Expert Systems With Applications 34(1), 256–265 (2008)
2. Huang, Y.-P., Chen, C.-H., Chang, Y.-T., Sandnes, F.E.: An Intelligent Strategy for Checking the Annual Inspection Status of Motorcycles Based on License Plate Recognition. Expert Systems With Applications 36(5), 9260–9267 (2009)
3. Kwasnicka, M., Wawrzyniak, B.: License Plate Localization and Recognition in Camera Pictures. In: Proc. of the 3rd Symposium on Methods of Artificial Intelligence, pp. 243–246 (2002)
4. Nijhuis, J.A.G., ter Brugge, M.H., Helmekolt, K.A., Pluim, J.P.W., Spaanenburg, L., Venema, R.S., Westenberg, M.A.: Car License Plate Recognition with Neural Networks and Fuzzy Logic. In: Proc. of the IEEE Int. Conf. on Neural Networks, ICNN 1995, vol. 5, pp. 2232–2236 (2002)
5. Kim, K.S., Kim, D.W., Kim, H.J.: A Recognition of Vehicle License Plate Using a Genetic Algorithm Based Segmentation. In: Proc. of the Int. Conf. on Image Processing, vol. 2, pp. 661–664 (1996)
6. Kim, K.K., Kim, K.I., Kim, J.B., Kim, H.J.: Learning-based Approach for License Plate Recognition. In: Proc. of the IEEE Signal Processing Society Workshop, vol. 2, pp. 614–623 (2000)
7. de Trazegnies, C., Urdiales, C., Bandera, A., Sandoval, F.: A Hidden Markov Model Object Recognition Technique for Incomplete and Distorted Corner Sequences. Image and Vision Computing 21(10), 879–889 (2003)
8. Dongliang, H., Feihu, Q., Jianfeng, L.: Recognition of Objects with Skew Distortion Based on Synergetics. Pattern Recognition Letters 20(3), 255–265 (1999)
9. Pan, X., Ye, X., Zhang, S.: A Hybrid Method for Robust Car Plate Character Recognition. Engineering Applications of Artificial Intelligence 18(1), 963–972 (2005)
10. Rahman, C.A., Badawy, W.: A Real Time Vehicle's License Plate Recognition System. In: Proc. of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS 2003), pp. 163–168 (2003)
11. Yu, M., Kim, Y.D.: An Approach to Korean License Plate Recognition Based on Vertical Edge Matching. In: IEEE Int. Conf. on Systems, Man, and Cybernetics, vol. 4, pp. 2975–2980 (2000)

12. Wu, C., On, L.C., Weng, C.H., Kuan, T.S., Ng, K.: A Macao License Plate Recognition System. In: Proc. of the 4th Int. Conf. on Machine Learning and Cybernetics, vol. 7, pp. 4506–4510 (2005)
13. Juntanasub, R., Spureerattanan, N.: Car License Plate Recognition Through Hausdorff Distance Technique. In: Proc. of the 17th IEEE Int. Conf. on Tools with Artificial Intelligence, ICTAI 2005, pp. 647–651 (2005)
14. Rauber, T.W., Steiger-Garcao, A.S.: 2-D Form Descriptors Based on a Normalized Parametric Polar Transform (UNL Transform). In: Proc. of MVA 1992 IAPR Workshop on Machine Vision Applications (1992)
15. Rauber, T.W.: Two-dimensional Shape Description. Technical Report: GR UNINOVA-RT-10-94, Universidade Nova de Lisboa (1994)
16. Frejlichowski, D.: The Point Distance Histogram for Analysis of Erythrocyte Shapes. Polish Journal of Environmental Studies 16(5b), 261–264 (2007)
17. Luengo-Oroz, M.A., Angulo, J., Flandrin, G., Klossa, J.: Mathematical Morphology in Polar-Logarithmic Coordinates. In: Marques, J.S., Pérez de la Blanca, N., Pina, P. (eds.) IbPRIA 2005. LNCS, vol. 3523, pp. 199–206. Springer, Heidelberg (2005)
18. Kukharev, G.: Digital Image Processing and Analysis (in Polish). Szczecin University of Technology Press (1998)
19. Zhang, D.S., Lu, G.: Generic Fourier Descriptor for Shape-based Image Retrieval. In: Proceedings of IEEE International Conference on Multimedia and Expo (ICME 2002), Lausanne, Switzerland, vol. 1, pp. 425–428 (2002)

Single Frame Rate-Quantization Model for MPEG-4 AVC/H.264 Video Encoders

Tomasz Grajek and Marek Domański*

Poznań University of Technology
Chair of Multimedia Telecommunications and Microelectronics
ul. Polanka 3, 60-965 Poznań, Poland
{tgrajek,domanski}@et.put.poznan.pl

Abstract. The paper describes a new model of MPEG-4 AVC/H.264 encoders. The model is proposed as a formula that expresses the number of bits per frame against the quantization step. Proposed are variants of the model for I-, P- and B-frames as well as for 3 intervals of quantization step values. The model has been derived from an extensive set of experimental data obtained by encoding numerous standard test video sequences. Experimental tests have proved good accuracy of the model proposed.

Keywords: compression, MPEG-4 AVC/H.264 standard, video encoding, video encoder modeling.

1 Introduction

The paper deals with modeling of advanced video encoders that are compliant with MPEG-4 AVC/H.264 standard [1]. In particular, the proposed models refer to encoders with compression performance being the same as that of the reference implementation of a video encoder of the standard MPEG-4 AVC/H.264 [13].

Modeling of video encoders is not only important for general research but it is also closely related to video encoder control. Despite of huge number of MPEG-4 AVC/H.264 video encoders working worldwide, designing efficient control algorithms is still a challenge. The basic parameter that is used to control an encoder is the quantization parameter QP that defines the quantization step size Q for transform coefficients. A typical goal of adjusting the quantization step size Q is to match the available channel bitrate. The Q parameter may be adjusted on the frame level, the slice level as well as on the macroblock level. Adjusting of Q value allows to control bitrate in a very wide range, e.g. from tens of megabits to tens of kilobits per second for standard definition video (i.e. 4CIF sequences).

In the informative part of MPEG-4 AVC/H.264 standard, a mode selection algorithm is described that exploits Rate-Distortion Optimization (RDO) [7]. At its input, this optimization algorithm needs a value of quantization parameter QP (directly related to Q), but this value can be determined from the mean

* This work was supported by the public funds as a research project.

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 384–391, 2010.

absolute difference (MAD) between the original image and the predicted one. As the MAD value is unknown until an image is encoded, a linear model for MAD prediction has been proposed [7]. Then quadratic rate-quantization model is used to calculate QP value [2]. In fact, many existing rate control schemes use the above described approaches.

Different approach to rate control was presented by He and Mitra in [5]. They have proposed to use a linear ρ-domain source model, where ρ denotes percentage of zeros in quantized transform coefficients. This model has proved to be very accurate in source content estimation, therefore several new rate control models have been developed as its variants, e.g. [8,9]. Unfortunately, direct implementation of this model inside MPEG-4 AVC/H.264 encoders is quite complicated [4,9].

The above mentioned models describe relationship between Q values and texture bits only i.e. bits for quantized transform coefficients. However, so called non-textured bits constitute a substantial portion of the total number of bits in AVC/H.264 bitstream. This number of bits should not be neglected, especially for low bitrates (see Fig. 1) [10,11]. Consequently, the encoder models should take both groups of bits into account. One solution is to apply separate models to texture and non-texture bits, at the price of higher complexity [6]. Another way is to develop a model for the total number of bits [10,12]. In this paper we exploit this second approach.

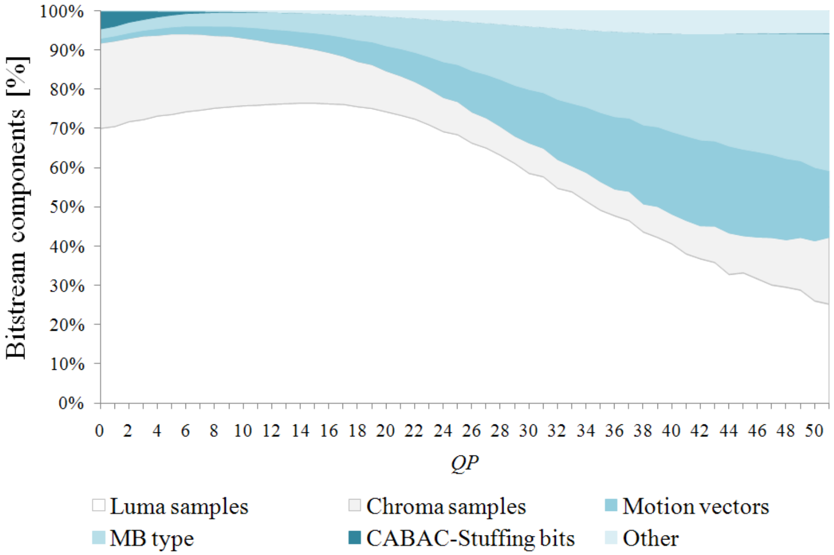


Fig. 1. Bitstream components for the reference MPEG-4 AVC/H.264 encoder averaged for all test sequences (encoder configuration: main profile, group of pictures: IBBPBBPBBPBBP, CABAC, deblocking filter switched on)

Our proposal is to treat a coder as a “black box” with one input (video sequence) and one output (bitstream), controlled by only one parameter Q, and to find an appropriate mathematical description that allows to estimate the total number of bits B with high accuracy. Here, number of bits B is the number of bits in the representation of a single individual frame. Mostly, this number is quite similar for frames of the same type taken from the same shot, i.e. I-, P- or B-frames taken from the interval between two scene cuts.

The model considered can be expressed as

$$B = f(Q, \Phi), \tag{1}$$

where Φ is a respective vector of parameters that depend on current video content and are obtainable by experimental data analysis.

In earlier papers (e.g. [2]), the authors have already proposed formulas for average number of all bits per frame within a single shot as well as for texture bits only in individual frames. In contrary to the above mentioned papers, here, we are going to derive a formula for all bits in individual frames.

2 Proposed Model

Our objective is to find the relationship between the number of bits B and quantization step size Q for a given frame type in a given sequence (Eq. 1). The relation will be established by analysis of experimental data.

Unfortunately, derivation of the model for the whole allowed range of Q values turned out to be inefficient (due to large approximation errors). Therefore, the entire allowed range of Q values has been divided into 3 intervals (Fig. 2, Table 1), and the model has been derived individually in each interval.

The most interesting central interval (Interval 2) covers Q values corresponding to bitrates from about 1 to 3 Mb/s (Fig. 2). In practical applications, it is the most useful range of bitrates for standard-definition video, i.e. 4CIF video sequences. Each test video sequence has a different range of Q values corresponding to the above mentioned range of bitrates. Therefore, for Interval 2, its low end has been set as the minimum value of Q that corresponds to 1 Mb/s in all test video sequences (see Section 3). Similarly the high end of Interval 2 has been set as the maximum value of Q corresponding to 3 Mb/s.

Table 1. Intervals for quantization step size Q and quantization parameter QP

Interval	Range of Q	Range of QP
1	$<0.625; 4.5>$	$<0; 17>$
2	$<5; 64>$	$<18; 40>$
3	$<72; 224>$	$<41; 51>$

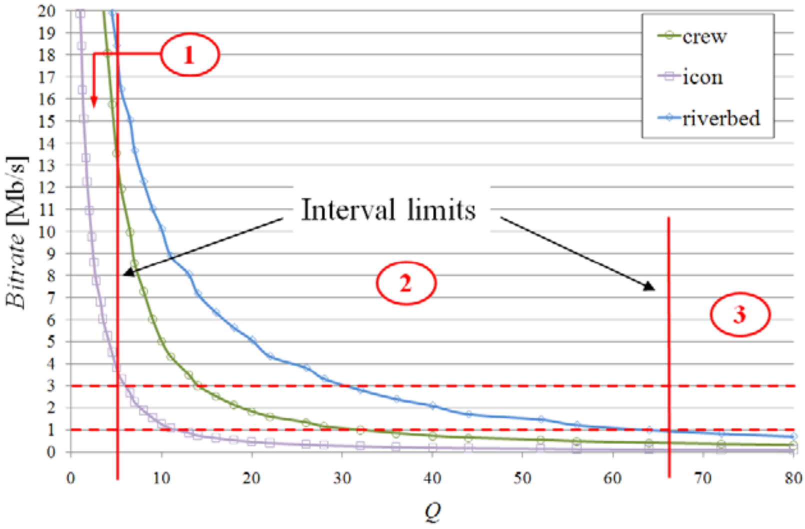


Fig. 2. Interval limits and experimental curves for 3 test sequences: *crew*, *icon*, *riverbed*. For the sake of clarity, range of Q and *Bitrate* have been clipped to 80 and 20 respectively.

2.1 Interval 2

For Interval 2, a hyperbolic model is proposed for number of bits per frame B

$$B(Q, \Phi) = \frac{a}{Q^b + c}, \tag{2}$$

where Φ=[a b c] is the vector of parameters that depend on sequence content. For these parameters, their values may change significantly at scene cuts. Within a shot, their values remain similar for all frames of a given type.

2.2 Interval 1 and Interval 3

For Interval 1 as well as for Interval 3, a cubic model is proposed,

$$B(Q, \Phi) = (a_1 \cdot d + a_2) \cdot Q^3 + (b_1 \cdot d + b_2) \cdot Q^2 + (c_1 \cdot d + c_2) \cdot Q + d, \tag{3}$$

where Φ=[d] is the vector with one parameter that depends on video content. The parameters a_1, a_2, b_1, b_2, c_1 and c_2 are the model constants that exhibits different values for both intervals and for each frame type (see Table 2). The model constants have been estimated experimentally using a set of test sequences.

3 Derivation of the Model

The proposed model has been derived individually for each interval of Q and each frame type. Function fitting was applied to experimental data collected for a set of 21 test video sequences. In that way function type was derived.

Table 2. Values of universal model constants for Intervals 1 and 3

Constants	Interval 1			Interval 3		
	I-frame	P-frame	B-frame	I-frame	P-frame	B-frame
a_1	$-1.65{\cdot}10^{-2}$	$-2.39{\cdot}10^{-2}$	$-1.85{\cdot}10^{-2}$	$-7.06{\cdot}10^{-8}$	$-1.25{\cdot}10^{-7}$	$-1.49{\cdot}10^{-7}$
a_2	$4.58{\cdot}10^{3}$	$1.49{\cdot}10^{4}$	$3.96{\cdot}10^{3}$	$-2.11{\cdot}10^{-3}$	$8.50{\cdot}10^{-4}$	$5.08{\cdot}10^{-4}$
b_1	$1.66{\cdot}10^{-1}$	$2.20{\cdot}10^{-1}$	$1.73{\cdot}10^{-1}$	$4.58{\cdot}10^{-5}$	$7.02{\cdot}10^{-5}$	$8.11{\cdot}10^{-5}$
b_2	$-2.63{\cdot}10^{4}$	$-8.67{\cdot}10^{4}$	$1.07{\cdot}10^{4}$	$8.90{\cdot}10^{-1}$	$-4.22{\cdot}10^{-1}$	$-2.35{\cdot}10^{-1}$
c_1	$-5.79{\cdot}10^{-1}$	$-6.82{\cdot}10^{-1}$	$-5.70{\cdot}10^{-1}$	$-1.07{\cdot}10^{-2}$	$-1.36{\cdot}10^{-2}$	$-1.51{\cdot}10^{-2}$
c_2	$-2.94{\cdot}10^{4}$	$3.21{\cdot}10^{4}$	$-2.08{\cdot}10^{5}$	$-8.46{\cdot}10^{1}$	$-6.21{\cdot}10^{1}$	$-2.97{\cdot}10^{1}$

For the same set of 21 test video sequences, the values of model constants (a_1, a_2, b_1, b_2, c_1, c_2) have been estimated individually in Interval 1 and Interval 3. These constants have been estimated by minimizing the maximum relative approximation error over the respective interval of Q values

$$\min_{\varPhi} \max_{Q} \epsilon(Q, \varPhi), \tag{4}$$

$$\epsilon(Q, \varPhi) = \frac{B_x(Q) - B(Q, \varPhi)}{B_x(Q)} \cdot 100\%, \tag{5}$$

where $B_x(Q)$ denotes the measured number of bits per frame and $B(Q, \varPhi)$ denotes the value calculated from the model.

For both, model type choice and model constant estimation, experimental data have been collected for 21 various 4CIF at 25Hz and 30Hz sequences. Each test video sequence consisted of 199 frames. The test video sequences had different motion characteristics. The following sequences have been used: *basket*, *bluesky*, *bus*, *cheer*, *city*, *crew*, *flow*, *football*, *harbour*, *ice*, *icon*, *pedestrian*, *riverbed*, *rushhour*, *soccer*, *station2*, *stefan*, *sunflower*, *tractor*, *universal* and *warner*. All sequences have been encoded using MPEG-4 AVC/H.264 reference software version JM_13.2 [13] (main profile, CABAC and RDO enabled, GOP: IBBPBBPBBPBBP). Each sequence has been encoded with all eligible values of Q. Of course, each value of Q was set indirectly by properly setting the value of quantization parameter QP. Sequences *bluesky*, *pedestrian*, *riverbed*, *rushhour*, *station2*, *sunflower* and *tractor* have been cropped to 4CIF resolution from their original 720p format.

4 Model Accuracy

In order to measure the accuracy of the model, the mean relative approximation error has been calculated separately for each encoded frame and for each interval of Q values. The experiments with additional test video sequences have shown that the model constants are chosen correctly, i.e. model accuracy remains similar also for the test material from outside of the initial set of 21 test sequences.

Table 3. Mean (i.e. averaged over all Q values) relative error for Interval 1

Model	Relative error averaged for all Q values for a given frame [%]			
	max over all frames	min over all frames	mean over all frames	std. deviation over all frames
I-frame				
cubic(4)	2.27	0.80	1.52	0.36
cubic(1)	70.02	1.16	3.84	7.98
P-frame				
cubic(4)	9.11	1.77	3.67	1.00
cubic(1)	152.16	2.79	7.93	14.50
B-frame				
cubic(4)	11.59	1.60	3.29	1.39
cubic(1)	197.15	2.43	9.29	20.00

Interval 1

In Table 3, the values of relative approximation error have been shown for Interval 1. The number in brackets indicates how many parameters the model has, e.g. "cubic (4)" refers to a version of the model from Eq. 3 with 4 content-dependent parameters

$$B(Q, \Phi) = a \cdot Q^3 + b \cdot Q^2 + c \cdot Q + d \tag{6}$$

Here, a, b, c and d are video-content-dependent parameters.

The average relative approximation error for the simplified cubic model (with one content dependent parameter d) is reasonable.

Interval 2

The characteristic parameters of the mean relative approximation error calculated individually for I-, P- and B-frames have been presented in Table 4.

The average relative approximation error is about 1.3%, 3.1% and 7.6% for I-, P- and B-frames, respectively. This proves usefulness and very good accuracy of the proposed model for encoders in the mostly used range of bitrates and quantization steps.

Table 4. Mean (i.e. averaged over all Q values) relative error for Interval 2

Frame type	Relative error averaged for all Q values for a given frame [%]			
	max over all frames	min over all frames	mean over all frames	std. deviation over all frames
I	7.36	0.81	1.30	0.55
P	35.13	0.92	3.11	3.03
B	55.12	0.89	7.60	6.77

Interval 3

For simplified cubic models, for P- and B-frames, model accuracy is poor in Interval 3 (Table 5), in contrary to the two previous intervals. In Interval 3, estimation of the number of bits is more difficult (especially for P- and B-frames) because for large quantization steps, numbers of bits are quite low. Therefore even small absolute errors in the number of bits, yield large relative approximation errors, especially for larger values of Q. In average, the approximation errors are probably acceptable but there exist some frames with really huge approximation errors.

Table 5. Mean (i.e. averaged over all Q values) relative error for Interval 3

Model	Relative error averaged for all Q values for a given frame [%]			
	max over all frames	min over all frames	mean over all frames	std. deviation over all frames
I-frame				
cubic(4)	4.40	0.73	1.62	0.49
cubic(1)	38.80	1.15	6.06	4.97
P-frame				
cubic(4)	34.52	0.88	3.20	2.57
cubic(1)	380.36	1.57	14.70	31.38
B-frame				
cubic(4)	42.25	0.00	6.54	4.62
cubic(1)	164.15	2.78	38.32	35.64

5 Conclusions

In this paper, a new model for MPEG-4 AVC/H.264 encoders has been described. This model is given as a function $B(Q)$, where B is the total number of bits in a frame. This is the difference with respect to other model described in the references that have proposed models for transform coefficient bits.

The new model is applicable for bitrate control. Moreover, it can be considered as a mathematical model of a video encoder that may be used in studies in video compression.

In this paper, the model parameters have been estimated for standard definition (SD) video. Nevertheless, similar relations hold for high definition (HD) video as well. The new model has been obtained by curve fitting that minimized the maximum approximation error of experimental data. These experimental data were obtained from extensive experiments with 21 video test sequences. The analysis of this huge set of experimental data resulted in proposal of the function type that is able to fit well the experimental data in individual intervals of quantization step size, and for different picture types. For two intervals, only one model parameter is needed that depends on sequence content. For these intervals, other parameters have been estimated as universal constants.

This model can be used to set a value of the quantization parameter QP for a given number of bits for an I- P- or B-frames. Tests proved that the model mostly fits experimental data very well in a wide range of bitrates. For the most useful range of bitrates for 4CIF sequences (Interval 2), the relative approximation error is about 1.3% for I-frames, 3.1% for P-frames and 7.6% for B-frames. For Interval 1 and Interval 3, the relative approximation errors are higher when only one content-dependent parameter is used.

We conclude that the proposed model is very accurate for moderate bitrates, i.e. for bitrates that are mostly used for video broadcasting, e.g. in digital television.

References

1. ISO/IEC 14496-10 (MPEG-4 AVC) / ITU-T Rec. H.264: Advanced Video Coding for Generic Audiovisual Services (2007)
2. Chiang, T., Zhang, Y.-Q.: A New Rate Control Scheme Using Quadratic Rate Distortion Model. IEEE Trans. Circ. Sys. Video Techn. 7, 246–250 (1997)
3. Grajek, T., Domański, M.: A Global Model of AVC/H.264 Video Encoders. In: 27th Pic. Cod. Symp., PCS, Chicago, Illinois, USA (2009)
4. He, Z., Chen, T.: Linear Rate Control for JVT Video Coding. In: Int. Conf. Inf. Tech.: Res. Edu., ITRE, Newark, USA, pp. 65–68 (2003)
5. He, Z., Mitra, S.K.: Optimum Bit Allocation and Accurate Rate Control for Video Coding via ρ-Domain Source Modeling. IEEE Trans. Circ. Sys. Video Techn. 12(10), 840–849 (2002)
6. Kwon, D.-K., Shen, M.-Y., Kuo, C.-C.J.: Rate Control for H.264 Video with Enhanced Rate and Distortion Models. IEEE Trans. Circ. Sys. Video Techn. 17(5), 517–529 (2007)
7. Li, Z.G., Pan, F., Lim, K.P., Feng, G., Lin, X., Rahardja, S.: Adaptive Basic Unit Layer Rate Control for JVT, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, doc. JVT-G012r1, Pattaya, Thailand (2003)
8. Lim, S.C., Na, H.R., Lee, Y.L.: Rate Control Based on Linear Regression for H.264/MPEG-4 AVC, Sig. In: Proc.: Image Communication, vol. 22(1), pp. 39–58. Elsevier, Amsterdam (2007)
9. Lin, G., Zheng, S., Hu, J.: A Two-Stage ρ-Domain Rate Control Scheme for H.264 Encoder. In: IEEE Int. Conf. Mult. Expo., ICME, Hannover, Germany, pp. 713–716 (2008)
10. Liu, Y., Li, Z.G., Soh, Y.C.: A Novel Rate Control Scheme for Low Delay Video Communication of H. 264/AVC Standard, IEEE Trans. Circ. Sys. Video Techn. 17(1), 68–78 (2007)
11. Xu, J., He, Y.: A Novel Rate Control for H.264. In:IEEE Int. Symp. Circ. Sys., ISCAS, Vancouver, Canada, vol. 3, pp. 809–812 (2004)
12. Zhuo, L., Gao, X., Wang, Z., Feng, D.D., Shen, L.: A Novel Rate-Quality Model Based H.264/AVC Frame Layer Rate Control Method. In: 6th Int. Conf. Inf., Com. & Sig. Proc., ICICS, Singapore, pp. 1–5 (2007)
13. H.264/AVC Software Coordination, http://iphome.hhi.de/suehring/tml

A Robust Method for Nose Detection under Various Conditions

M. Hassaballah[1,2], Tomonori Kanazawa[3], Shinobu Ido[3], and Shun Ido[1]

[1] Department of Computer Science, Ehime University, 790-8577, Japan
[2] Department of Mathematics, Faculty of Science, South Valley University, Egypt
[3] eCompute Corporation, 2821-4, Matsuyama, Ehime, 791-8042, Japan
m.hassaballah@ic.cs.ehime-u.ac.jp, ido@cs.ehime-u.ac.jp

Abstract. In this paper, a robust fully automatic method for nose field detection under different imaging conditions is presented. It depends on the local appearance and shape of nose region characterized by edge information. Independent Components Analysis (ICA) is used to learn the appearance of nose. We show experimentally that using edge information for characterizing appearance and shape outperforms using intensity information. The influence of preprocessing step on the performance of the method is also examined. A subregion-based framework depending on statistical analysis of intensity information in the nose region is proposed to improve the efficiency of ICA. Experimental results show that the proposed method can accurately detect nose with an average detection rate of 95.5 % on 6778 images from six different databases without prior detection for other facial features, outperforming existing methods.

1 Introduction

Detection of facial features such as eye, mouth, and nose remains a vital and challenging problem as it plays an important role in face recognition systems and many other applications. Numerous works have been published on facial features detection, notably on eye and mouth detection. Although, nose is not less important than eye or mouth in recognizing a human face [1], does not affect so much by facial expressions and in several cases is the only facial feature which is clearly visible during the head motion [2], most of these works ignore or avoid to detect nose [3]. Even the other works which detect nose did not introduce mature methods for nose detection. Some of them detect nose depending basically on the prior detection of eye center and considering nose can be located within certain pixels below the line connecting the centers of two eyes [4], and other works detect nose depending on the reflection of light on the nose tip or using projection methods [5]. However, all existing projection methods do not consider complex conditions such as illumination and pose variation, thus they will fail under these imaging conditions.

Few approaches have been developed in the last five years to detect nose. In [6], nose is detected based on both gray-level and color information. Gizatdinova and Surakka [7] introduce a method for facial landmarks detection based on

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 392–400, 2010.

extracting oriented edges and constructing edge maps at two resolution levels. They report that the method can achieve average nose detection of 78% on 330 images from Pictures of Facial Affect database. The method was not fully automatic and required manual classification of the located edge regions. On the other hand, 3D images are used in [8] and [9] to locate nose tip and nose ridge. These methods calculate the neighboring effective energy of each pixel to locate suitable nose candidates, then calculate the neighboring mean and variance of each pixel and use SVM to further narrow down the nose tip candidates. Finally, the nose tip is found by choosing the area which has the top three densest nose candidate regions. Unfortunately, none of these methods report an exact nose detection rate over a common database.

In this paper, a simple method based on ICA and analysis of intensity information is presented to detect nose field. The local edge information of nose image in horizontal direction is used to characterize nose appearance and shape, which appears to give significant higher performance than the intensity information with variance normalization preprocessing. The detected nose field is quite adequate to be used for face recognition or as a search area for subsequence exact nose tip and nostrils detection. The proposed method is evaluated on XM2VTS, BioID,1500 images of FERET, JAFFE, PICS, and MIT-CBCL databases, which stressing real world conditions.

2 Overview of the Method

2.1 Computing ICA Basis Images

The FastICA algorithm [10] is implemented to estimate the ICA basis images as it has desirable properties compared to the existing ICA algorithms. To compute the ICA basis vectors, a set of 200 nose images from different individuals is used in the training step, each one is of size w x h pixels. The width w and height h of the training images are determined experimentally with respect to the face width W_F.

Edge or gradient structure is very characteristic of local object appearance and shape [11]. The edge information of nose training images in horizontal direction is calculated using Sobel and Prewitt operators without smoothing. For Prewitt operator, the training image is convoluted with the mask [-1,0,1]. It is noted that vertical edge and magnitude were also tried, but in experiments they were shown to be inferior. This may be due to the fact that nose has two nostrils which can often be characterized rather well by the horizontal distribution of local intensity gradients. To minimize the impact of noise or variations in illumination, two preprocessing techniques are examined; norm normalization and variance normalization. Then, the FastICA algorithm is run to compute basis images from the three preprocessed training images; gray intensity, horizontal Sobel, and horizontal Prewitt. Figure 1 shows example of 30 ICA basis images computed from sobel training images with variance normalization. It is clear that the computed basis images are spatially localized to nose features such as nostrils and edges. This localization might robust the performance to local distortion.

2.2 Nose Detection Using ICA

Searching for nose in the image can be done either directly in the entire image, or rely on the output of a face detector indicating that nose is present in the image. Unfortunately, searching in the whole image is not a suitable method for real time implementations and is more prone to errors. Therefore, a face detector [12] is applied first to locate the facial region, then nose searching is done in the located region. The located face region is normalized to a fixed size of W_F x W_F pixels. The normalized face is scanned to find the region which might contain nose, the regions data are treated as w x h dimensional vectors. Each region (vector) is locally preprocessed using norm or variance normalization according to the preprocess used in the training step. The subspaces corresponding to the feature point vector in the w x h dimensional feature space can be expressed as linear subspaces spanned by multiple ICA vectors. The projection angle θ of an input vector (nose candidate region) projected onto the ICA subspace represents the extent to which the input vector is analogous to the feature point vector. For verification the value of θ; specifically $Cos\,(\theta)$ between the input vector and each feature point's subspace (ICA basis vectors) is obtained. This angle measure is invariant under linear changes in the contrast of the image and furthermore the cosine similarity measure was previously found to be effective for face processing [13]. The nose field is the input vector that falls into the subspace with the smallest θ or highest similarity. In other words, the similarity $S_{\theta_j}(R_j, ICAvectors)$ between the region R_j, $j = 1, 2, ..., M$ (number of regions) and ICA basis vectors is calculated using $Cos\,(\theta)$ of the projection component

$$S_{\theta_j} = Cos^2\theta_j = \frac{\sum_{i=1}^{n} \langle V_j, BaseVector_i \rangle^2}{\|V_j\|^2} \tag{1}$$

where n is the number of selected basis vectors that form the subspace and $\langle V_j, BaseVector_i \rangle$ is the inner product between input vector (representing j-th

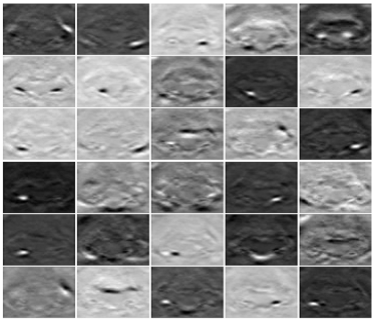


Fig. 1. The first highest 30 ICA basis images

nose candidate region) and i-th base vector (Fig. 1). The region R_k with the highest similarity S_{θ_k} is the nose field, i.e.

$$S_{\theta_k} = arg \quad max\{S_{\theta_1}, S_{\theta_2}, S_{\theta_3}, ...S_{\theta_M}\} \tag{2}$$

For the experiments in this work, a face of size 200 x 200 (*i.e.* W_F=200) pixels is used, thus the size of the nose region is set to be 70 x 50 pixels.

2.3 Subregion-Based Scheme (ICA+NP)

The performance of the method described in section 2.2 is shown in Table 1 (i.e. **ICA**). This performance can be improved further using statistical nose properties (**NP**). Experimentally, we note that in some cases ICA is able to give a close estimate around the position of nose location and it may fail to provide accurate detection of its location, because shadows, facial expression and other kinds of noise can make neighboring regions look like a nose. In these conditions the similarity between an input vector (nose candidate region) and ICA basis images will be high. Figure 2 shows the ten highest similarity value regions, most of them are gathered around nose. How can one choose the best region, which represents the accurate nose field from these 10 regions? To answer this question, analysis of the nose region is required to find its unique features.

(i) Similarity of both sides: The left and right sides of nose are similar in a front-view face as shown in Fig. 3(a), this property of similarity can be measured using Euclidean distance between both sides as follows.

$$S(L, R) = \sqrt{\sum_{w/2} (P_L - P_R)^2} \tag{3}$$

where P_L and P_R are the pixel gray value in left and right side respectively. It is also noted that the similarity between both sides in the lower part of nose is clearer than the upper part which looks like skin. This fact should be taken in account at calculation of the similarity between both sides of human nose. Thus, the relative contribution of the similarity between both sides in the upper and lower part is controlled by α where $0 < \alpha < 1$. Hence, the total Euclidean distance between left and right side of nose is calculated based on

$$E(L, R) = \alpha S_{upper} + (1 - \alpha) S_{lower} \tag{4}$$

where S_{upper} and S_{lower} are calculated using (3) over upper and lower half of nose region respectively, the smaller the distance $E(L, R)$ the higher similarity. The control α is set to be 0.4 based on analysis of the used training data which means that the similarity between both sides affects by 40% in the upper part and 60% in the lower part of total similarity between left and right sides of nose.

(ii) Dark-White-Dark (DWD) property: Also, the lower part of nose is characterized by two dark nostrils and a light subregion due to the reflection of light on the nose as shown in Fig. 3(b). This property can be identified by the average of gray intensity in each subregion, where the average in the two nostrils regions is less than the average of middle lighter subregion containing nose tip.

(iii) The variation in lower/upper parts property: On the other hand, when the face is rotated some degrees these two properties are despaired and the only clear property is the variation between lower part and upper part as shown in Fig. 3(c). This variation can be measured by the variance in each part, where the variance on a domain Ω of a nose image $I(x, y)$ can be defined as

$$V_\Omega = \frac{1}{A_\Omega} \sum_{(x,y)\in\Omega} [I(x, y) - \bar{I}_\Omega]^2 \tag{5}$$

where A_Ω and $\bar{I}_\Omega$ are the area and the average gray intensity on the domain Ω, respectively. The variance of lower part should be greater than that of the upper part. Based on this analysis, we search for a certain region among the ten highest regions detected by ICA (shown in Fig. 2) which satisfies the properties(i)-(iii).

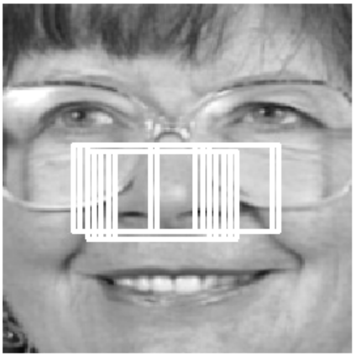


Fig. 2. The ten highest similarity value regions detected by ICA

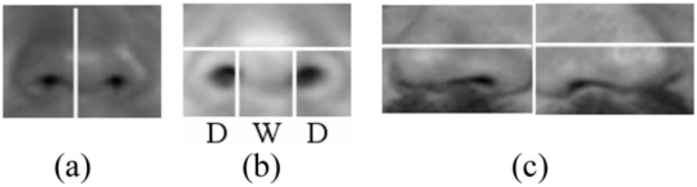
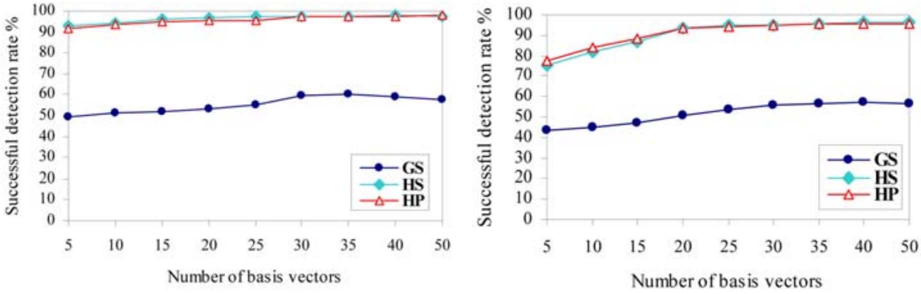


Fig. 3. Analysis of intensity information in the nose region

3 Experimental Results

To evaluate the performance of the proposed method, various experiments on common databases stressing different views, illumination conditions and facial expression have been carried out. The first experiment has been carried out on XM2VTS database to find optimum number of ICA basis images and to investigate the impact of preprocessing. As mentioned previously the training images are provided for FastICA algorithm as Gray Scale (GS), Horizontal Sobel (HS), and Horizontal Prewitt (HP). The performance of each representation with norm and variance normalization preprocess at different number of ICA basis images is shown in Fig. 4, which shows that using edge information to describe the appearance of nose outperforms intensity information. Also, the performance with variance normalization (Fig. 4(a)) is higher than with norm normalization (Fig. 4(b)). The highest detection rate for both processing methods achieved when the number of ICA components is in the range of 30 to 40 basis images and after that there is only a marginal improvement in the detection rate. It should be noted that the performance of the three representations gets worse without preprocessing where the highest detection rate was 60.5% achieved by GS. These results indicate that preprocessing is vital, and without it the results would be even worse. Looking closely, one can see that Sobel operator is slightly better than Prewitt operator at most ICA dimension. From this experiment we can concluded that Sobel operator achieved 97.4% detection rate outperforming Prewitt operator which achieved 97.2% at 30 ICA basis images with variance normalization (Fig. 4(a)). Also, there is trade-off between running time and ICA dimension. Using fewer basis images requires less storage and results in faster matching, hence 30 basis images are sufficient.



(a) Variance normalization preprocessing

(b) Norm normalization preprocessing

Fig. 4. Performance versus ICA dimension and preprocessing on XM2VTS database

In the second experiment, nose field is detected based on the highest response to ICA basis images computed from edge information (HS) with variance normalization. The number of ICA basis vectors is set to 30 vectors. The results of this experiment are reported in Table 1 (**ICA**). Because of noise in the images due to lighting conditions or facial expressions, the detected region is not accurate and hence the whole successful detection rate (i.e. images with correctly detected nose field relative to the whole set of facial database) is not so high, especially on BioID and JAFEE databases. The best results have been obtained on the XM2VTS database because the image quality is very high, the illumination is always uniform, and the image scale is fixed.

In the third experiment, we search for nose field among the most similar 10 regions to nose with high response to ICA basis images. The region that satisfies the three local nose properties mentioned previously is the searched nose field. The results reported in Table 1 (**ICA+NP**) show that the suggested subregion-based framework significantly improves the performance of ICA in most databases, while the average detection rate improved from 93.9% to 95.5%. It is well known that the computational complexity of (3) and (5) is relatively high. But this does not affect on the computation time of the proposed method, because they are calculated only for the ten regions and on small subregions of these regions. The average execution time of the method on a PC with Pentium III 1.0 GHz and 256Mb RAM is less than 100 *msec*. Compared with other nose detection methods [7] and [14] achieved average detection rate of of 78% on 330 images and 95.3% on 963 images respectively, the proposed method is capable of detecting nose with average detection rate of 95.5% on 6778 images. Besides the high detection rate achieved by the method, nose detection does not require any prior detection for other facial features such as eyes or mouth as in these methods.

Table 1. Detection rates of the proposed method

Database **Method**	**XM2VTS**	**BioID**	**FERET**	**JAFFE**	**PICS**	**MIT-CBCL**	**Average** **Rate**
ICA	95.8%	91.8%	94.3%	91.5%	93.2%	94.4%	93.9%
ICA+NP	99%	92.3%	96%	96.2%	95%	95.7%	95.5%

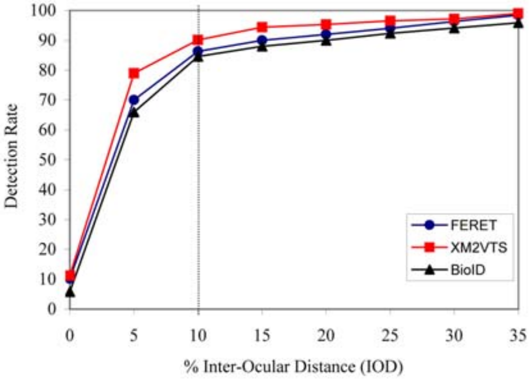


Fig. 5. Performance of nose tip detection

On the other hand, in the case of considering the center point of the located region is the nose tip, the performance of nose tip detection can be evaluated with respect to inter-ocular distance (IOD) (i.e. distance between eyes centers) by computing Euclidean distance (in terms of pixels) between the estimated point

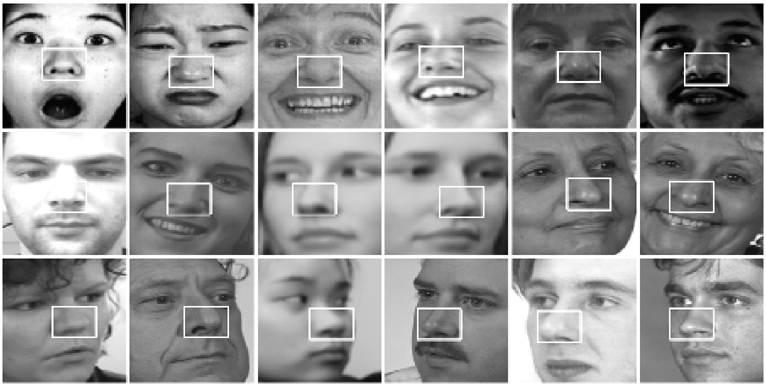


Fig. 6. Examples of successful nose detection under various conditions

and its ground truth. Figure 5 shows the performance of nose tip detection on three of databases FERET, BioID, and XM2VTS, which provide ground truth of nose tip. The detection rate is not less than 83% at the acceptable threshold 10% of IOD. Examples of successful nose detection are shown in Fig. 6.

4 Conclusion

This paper introduced a method to detect the nose field from gray scale facial images. The method detects nose based on the highest response to basis images. Independent Components Analysis (ICA) is used to compute basis images as it provides a better probabilistic model of training data through decorrelating higher-order statistics in addition to the second-order moments. The appearance of nose is characterized by edge information, which gives better performance than intensity information. The importance of preprocessing to the method is investigated. In order to further improve the performance of the method, we proposed a subregion-based framework that depends on statistical analysis of intensity information in the nose region. The efficiency of the method was assessed in different databases stressing various conditions. It has been shown by experimental results that the proposed method can accurately detect nose with a higher detection rate compared to the state of the art methods.

Acknowledgement

The authors would like to thank the owners of all databases used in the experiments of this work and the Egyptian government for supporting this work under Grant No.1/13/101-2006-2007.

References

1. Queirolo, C., Silva, L., Bellon, O., Segundo, M.: 3D face recognition using simulated annealing and the surface interpenetration measure. IEEE Trans. on Pattern Analysis and Machine Intelligence 32(2), 206–219 (2010)
2. Gorodnichy, D., Roth, G.: Nouse 'use your nose as a mouse' perceptual vision technology for hands-free games and interfaces. Image and Vision Computing 22, 931–942 (2004)
3. Asteriadis, S., Nikolaidis, N., Pitas, I.: Facial feature detection using distance vector fields. Pattern Recognition 42, 1388–1398 (2009)
4. Sankaran, P., Gundimada, S., Tompkins, R.C., Asari, V.K.: Pose angle determination by faces, eyes and nose localization. In: IEEE CVPR 2005 (2005)
5. Bevilacqua, V., Ciccimarra, A., Leone, I., Mastronardi, G.: Automatic facial feature points detection. In: Huang, D.-S., Wunsch II, D.C., Levine, D.S., Jo, K.-H. (eds.) ICIC 2008. LNCS (LNAI), vol. 5227, pp. 1142–1149. Springer, Heidelberg (2008)
6. Campadelli, P., Lanzarotti, R.: Fiducial point localization in color images of face foregrounds. Image and Vision Computing 22, 863–872 (2004)
7. Gizatdinova, Y., Surakka, V.: Feature-based detection of facial landmarks from neutral and expressive facial images. IEEE Trans. on Pattern Analysis and Machine Intelligence 28(1), 135–139 (2006)
8. Chew, W.J., Seng, K.P., Ang, L.M.: Nose tip detection on a three-dimensional face range image invariant to head pose. In: Proc. of the Int. MultiConference of Engineers and Computer Scientists IMECS 2009, pp. 858–862 (2009)
9. Xu, C., Wang, Y., Tan, T., Quan, L.: Robust nose detection in 3D facial data using local characteristics. In: ICIP 2004, pp. 1995–1998 (2004)
10. Hyvarinen, A., Oja, E.: Independent component analysis: algorithms and applications. Neural Networks 13, 411–430 (2000)
11. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR 2005, pp. 886–893 (2005)
12. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision 57(2), 137–154 (2004)
13. Bartlett, M.S., Movellan, J.R., Sejnowski, T.J.: Face recognition by independent component analysis. IEEE Trans. on Neural Networks 13(6), 1450–1464 (2002)
14. Kim, K.A., Oh, S.Y., Choi, H.C.: Facial feature extraction using PCA and wavelet multi-resolution images. In: IEEE FGR 2004, pp. 439–444 (2004)

Segmentation of Moving Cells in Bright Field and Epi-Fluorescent Microscopic Image Sequences

Marcin Iwanowski[1] and Anna Korzyńska[2]

[1] Warsaw University of Technology, Institute of Control and Industrial Electronics
ul. Koszykowa 75, 00-662 Warszawa, Poland
iwanowski@ee.pw.edu.pl

[2] Nalecz Institute of Biocybernetics and Biomedical Engineering PAS
ul. Ks.Trojdena 4, 02-109 Warszawa, Poland
anna.korzynska@ibib.waw.pl

Abstract. The monitoring of the dynamics of stem cells' growth in culture is important in regenerative medicine. In this paper the method of cells' images segmentation based on alternating microscopic imaging with bright field (BF) and epifluorescent (EF) images is proposed. The method consists of two principal stages: coarse segmentation of EF images followed by fine segmentation on BF ones. The latter is based on the morphological watershed from markers produced in the first stage. Due to the fact that sequence of EF is shorter than BF one, markers cannot be produced directly for all BF images. In order to create them, an additional step of morphological interpolation of markers is applied.

1 Introduction

Monitoring and analysis of stem culture quantity is crucial for reliable optimization of cells' culturing methods. For therapeutic purposes it is needed to find effective methodology for donors cells' population increscent to the proper size according to the clinical procedures [20]. It is common to evaluate cells' cultures with human eye or using immunohistochemistry and flow cytometry [15]. The idea of cultures monitoring acquiring time sequences of images and extracting information from images using computer supported microscopy is not a new one [15,22,19,13,12,4,1]. In this paper, the extraction of stem cells based on synergic information coming from two types of imaging modalities is proposed.

The aim of the presented investigation is to develop a new method for cells extraction from static images in time sequence, which allows next to analyze cells' behavior (its movement, shape changes, proliferation and death). This method of image segmentation can use both types of information from BF image and from EF image, if both of them are available at the same time. According to other investigators, a segmentation method using only BF images is difficult, but some successful applications are reported [8,9,17,5,2]. Most of reported cases of successful segmentation is concerned with images of cytological smears. They are similar to adherent cells' culture images, because in both cases there is a layer of single cells on the microscopic slide. For images of cells' smear the investigators

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 401–410, 2010.

propose two types of the segmentation procedure [28,24]: (1) using a simple morphological operation, supported by the artificial intelligence methods, such as SVM or decision trees [11,16,6] or (2) using a classical threshold or classification algorithms [7,8].

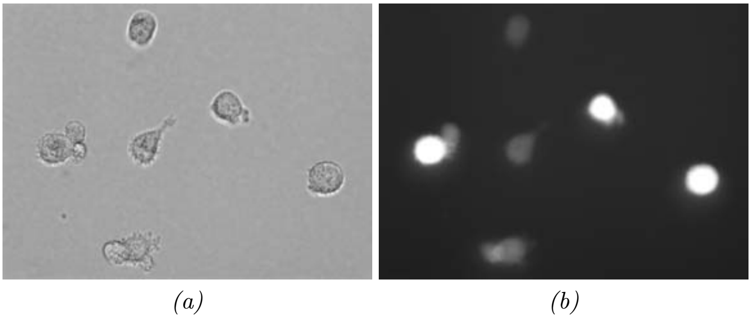


Fig. 1. Examples of images: *(a)* – bright field (BF) image, *(b)* – corresponding filtered epi-fluorescent (EF) image

In this investigation it is assumed that image sequences are acquired alternately using bright field (BF) and epifluorescent (EF) techniques for fixed image plane and for neural stem cells transfected with green fluorescent protein (*gfp*) [14]. BF images are collected with constant time increment that varies from 1 to 30 minutes. EF images are registered also in series, but with time increment, which is greater than in the BF-case because of the fact that fluorescence is exhausted with time what causes decrease of quality of EF images. So practically, a single EF-frame is captured with every k-th BF image, where $k > 1$ is set up manually. Rare capturing of EF images among more frequent BF imaging sequences allows increasing time of overall observations thanks to both: light used for fluorescence excitation (UV) and fluorescent agent itself (fluorescent protein products) are more destructive for living cells then conventional light used in BF microscopy [21].

The proposed method assumes that the input set of images consist of two sequences. The first one is the sequence of BF images, and the second one, which is k-times shorter, of EF images. Each EF sequence frame is registered just after corresponding BF frame with hardware minimal delay (about 25±1 second in our case). Cells have different appearance at both kinds of frames. In BF images cells are highly textures, what causes them difficult to segment. The same cells in EF images are characterized by uniform gray level, higher than the gray level of the background, which makes them relatively easy to segment using thresholding preceded by the appropriate filtering. There exist, however, an important disadvantage of such segmented images. The contour of cells presented on EF- frames is coarse and is not precisely following the real cell outline. This is due to the fact that EF-images register electromagnetic radiation light emission what introduces some additional blurring. Consequently, the best approach to perform the proper segmentation is a combination of coarse segmentation on EF-images followed by fine segmentation performed on appropriate BF frame. The latter is performed using the morphological watershed that makes use of markers obtained in the coarse segmentation. Such approach is possible only if there exist two frames: BF and EF showing a single area at the same time (with an accuracy of hardware delay). Unfortunately, as mentioned earlier, not every BF frame is accompanied by a EF image. This make the above outlined segmentation possible only for some frames. For the remaining ones, markers obtained by coarse segmentation of fluorescent image are missing because such frame itself is missing.

In order to recover the missing markers, the morphological interpolation of known ones is performed. This is a shape interpolation technique applied on binary images. Due to the fact that cells are moving, also this movement should be considered during the interpolation phase. It can be assumed that this movement between two consecutive known markers is linear. Thanks to that the linear translation should be combined with the morphological interpolation, which is a simplified case of morphological-affine interpolation [23].

2 Proposed Method

The proposed method is based on morphological image processing [27,18] and consists of three steps: coarse segmentaion of EF images, marker interpolation and fine segmentation. The diagram presenting processed images in case of one epifluorescent image per 4 bright-field ones is shown in Fig. 2. There are two sets of input images: bright-field (BF1,...,BF5), epifluorescent (EF1, EF2). Coarse segmentation of EF1 and EF2 images results in two marker images M1 and M5. Missing markers M2,M3 and M4 are morphologically interpolated based on M1 and M5. Finally, all 5 images M1,...,M5 are used as markers for watershed transform of gradient images produced from BF1,...,BF5. Watershed segmentation produces fine segmentation result marked in diagram as S1,...,S5.

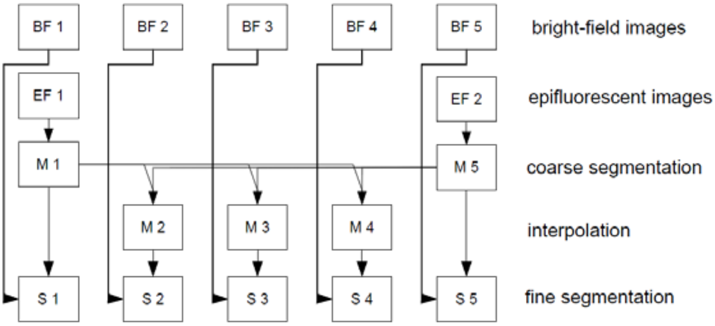


Fig. 2. Diagram of the proposed method

Particular phases of processing are described in details below.

2.1 Coarse Segmentation

Coarse segmentation is performed on the EF image. An example of such image is presented in Fig. 1*(b)*, a picture showing chosen single cell (region of interest - ROI) is shown in Fig. 3*(a)*. The EF images are characterized by relatively high level of noise and by non-uniform intensity of emission light distribution across on cell. In order to remove noise the morphological opening filter was used. The result of this filtering is presented in Fig. 3*(b)*. The filtered image is then thresholded. As a result the marker representing main cell body area is obtained (see Fig. 3*(c)*).

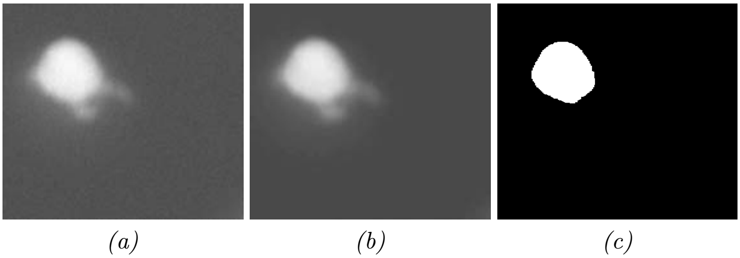


Fig. 3. Coarse segmentation from EF images: *(a)* the original EF image (region of interest - ROI) is *(b)* filtered by morphological opening, and *(c)* thresholded

In order to get the precise outline, the fine segmentation process based on the morphological watershed using the just obtained markers is applied. This process will be described in section 2.3.

The coarse segmentation result cannot be computed for each BF frame due to the fact that the EF frames are available only for some of BF ones. In order to create missing markers, which are necessary to compute the fine segmentation result, the morphological interpolation approach is used. This step is described in the next section.

2.2 Markers' Interpolation

Let assume that EF frames are available for every k-th BF ones. This means that, in order to get the complete sequence of markers, $k-1$ markers have to created between every two obtained using the procedure of coarse segmentation described in the previous section. The images with the markers are binary ones, where value 1 is assigned to all pixels belonging to the roughly estimated area of cell, while value 0 – to the background pixels. The morphological interpolation approach used here is based on computation of the interpolation function [26]. Due to the fact that the cells are moving, an variant of this approach that combines morphological interpolation with the affine transformation called morphological-affine interpolation [23] is used.

The transformation of markers between two consecutive frames might be considered as a combination of the modification of shape and the movement, that – in such relatively short time period – can be considered as linear. The latter implies a simple form of the affine transformation consisting only of translation along the straight line. This form of simplified description of cell movement is typically used in cells' motility investigations and is called a trajectory.

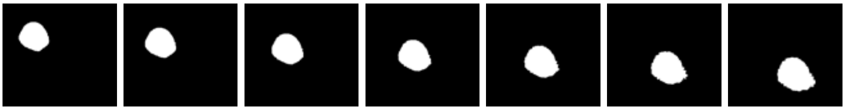


Fig. 4. Interpolation of markers – the first and the last are binary images showing results of the coarse segmentation and computed using EF images, while all the others are results of morphological interpolation using initial position from first and final position from the last image

The result of morphological-affine interpolation of a single cell marker is shown in Fig. 4. In this example the first and the last are binary images shows results of coarse segmentation done using the EF images. Intermediary images present results of the morphological interpolation.

2.3 Fine Segmentation

After the coarse segmentation and interpolation steps, two sequences have been obtained – the original sequence of BF images and the sequence of markers pointing at the cells. Thus each BF frame is equipped with the markers pointing at the cells. Thanks to that, the final segmentation of BF images by morphological watershed of the modified gradient image, is possible.

The input image for the watershed transformation must be prepared in order to get to correct segmentation result. At the beginning, the gradient is computed. On the original image each cell is outlined by the halo-effect. The thickness of the halo-effect outline would result in a double gradient-line in case of the classical approach (either linear or morphological) is used. In order to avoid this effect, the morphological thick-gradient [18,3] by erosion is used. The definition of this gradient is the same as of the usual case: subtraction the result of erosion from the original image. In the 'thick' case however, the structuring element used by the erosion is a large one. The result of gradient computation of the original image is shown in Fig. 5*(b)* (while original image is shown in Fig. 5*(a)*).

Quality of the gradient influences directly the quality of the final watershed. In particular, the gradient regional minima are strictly connected with the regions outlined by the watershed lines (catchment basins): each catchment basin refers to a single minimum. This property of the watershed transform causes often an oversegmentation when the numbers of regions on the segmentation result

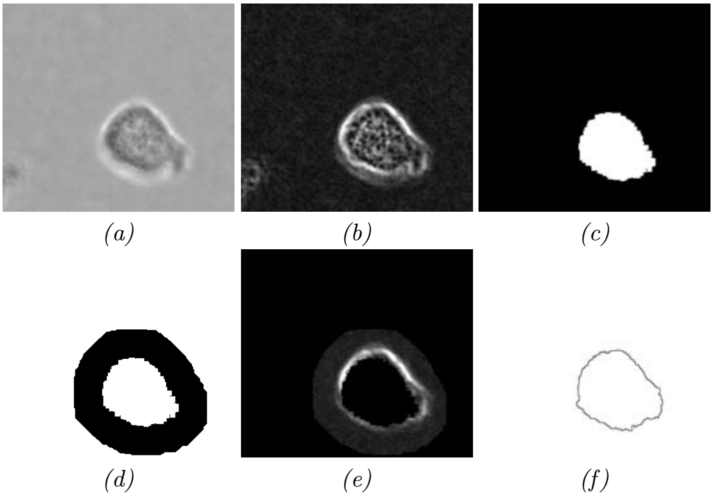


Fig. 5. Fine segmentation process:*(a)* original BF image, *(b)* its morphological thick-gradient by erosion, *(c)* marker, *(d)* inner and outer markers which forms belt to cell's edge detection, *(e)* minima imposed on gradient image, *(f)* the result of the watershed

exceeds the actual number of objects on the image to be segmented. In order to avoid the oversegmentation, the minima's imposition procedure [18,3] is used. The minima's imposition is a tool, which modifies gradient in order to force the presence of image regional minima (crucial for correct result of the watershed) in given image regions. This operation is based on morphological reconstruction and removes all the regional minima which are not marked by special markers. It produces thus new minima in all regions covered by markers. Consequently, it is a typical supplement of watershed transformation, where marker image is created separately in such a way that each marker points either at particular object present on the input image or at the background. Thus, two kind of markers can be defined: inner, which indicates objects (cells in our case) and outer that points at the background area. Minimas' imposition with such markers guarantees that the watershed line will be located inside the belt-like area between the inner and outer markers.

In the current study, the inner markers for the watershed are equivalent to the markers obtained either by the coarse segmentation or by the interpolation. As the outer markers, the complement of dilation of the inner ones is considered. Figure 5*(c)* presents the inner markers and Fig. 5*(d)* the union of inner and outer used in the minima's imposition step. The result of minima's imposition is shown in Fig. 5*(e)* as a belt rounding the cell. This belt is characterized by the fact that it indicates the area within, which the proper outline of the cell is located. The result of minima imposition makes such input image for the watershed which ensure results without oversegmentation. Image that is the result of the watershed (shown in Fig. 5*(f)*) contains the precise and unequivocal outline of the cell.

3 Results and Conclusions

3.1 Materials and Experiments

The Human Umbilical Cord Blood derived Neural Stem Cells (HUCB-NSC line) established and transfected *gfp* in NeuroRepair Dep. Lab; Mossakowski Medical Research Centre PAS were used to capture the cell images [20,14]. The stem cells were seeded onto plastic culture dishes at density 105 cell/ml in standard conditions (37°C, 5% CO_2 and humidity 95%) for 24h (for details see [4]). Then the images' sequences were acquired using Cells Behavior Monitoring System (CBMS; IBBE PAS), with CoolSNAP-Pro cf camera mounted on the Olympus IX70 inverted microscope (objective x40, 0.60 NA, WD 2.6) with shatters position controllers. The sequences of images were captured with basic time increment equal 10 min. and k=1, 2, 5 as monochrome 1392 x 1040 pixels, 12 bits deep gray scale images showing culture dishes area of 220μm x 165μm.

3.2 Results of Fine Segmentation

The result of the segmentation using the proposed method presented in figures throughout the paper was obtained from the input sequences: BF, consisting of 9 frames and EF, consisting of 4 frames. Thus each EF frame was acquired along with every second BF frame. In order to test and show the results of marker interpolation, only the first and the last EF frame were considered in the segmentation process, two intermediary was used to verify the results of marker interpolation. Parameters applied to consecutive operations of the method were the following. In the coarse segmentation phase: the structuring element used in opening filter – 9×9 square, threshold was set-up as 0.7 of the maximum gray value of the filtered ROI. In the marker interpolation phase: the number of frames to be interpolated was equal to $k = 7$. In the fine segmentation phase: the thick gradient was computed using the structuring element equal to disk of radius 9 and the structuring element of the dilation used to compute the outer markers was disk of radius 50.

The final result of segmentation of all the sequence frames is shown in Fig. 6.

3.3 Comparison of Interpolation Results with Results of Coarse Segmentation

Two intermediary EF frames were not used for the interpolation of markers. They were kept to evaluate quality of the marker interpolation. Both were roughly segmented and next compared to the appropriate interpolated markers. The comparison results are shown in Fig. 7, each EF number of the frame refers to appropriate number of BF frame. On pos.*(b)* and *(c)* both interpolated and segmented markers are presented. Interpolated markers are slightly greater (dark-gray outline) than the segmented ones. Mean square errors computed within the images are equal to 0.0095 (3th frame) and 0.0149 (5th frame). This, relatively small difference in size do not influence usefulness of the markers for the fine segmentation, as it was shown in Fig. 6.

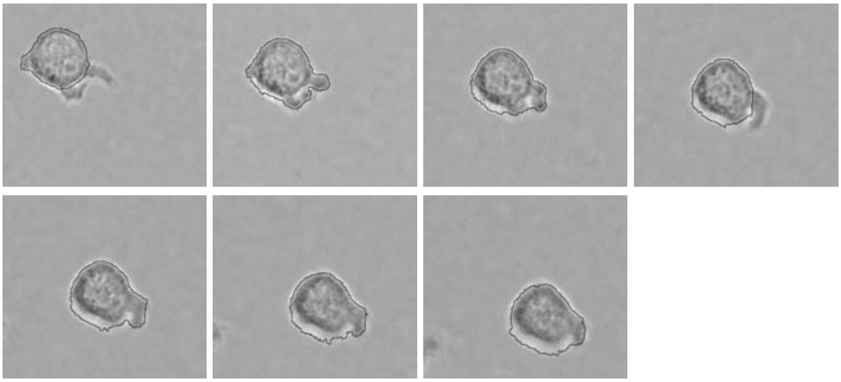


Fig. 6. The results of segmentation of chosen single cell from testing sequence. This results obtained using markers calculated on the basis of EF images for the first and the last image while for middle 5 images markers are interpolated (markers in Fig. 4).

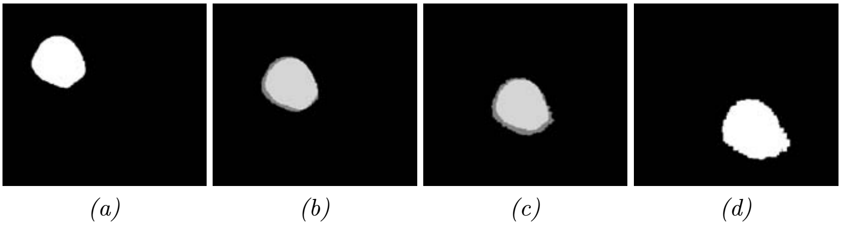


Fig. 7. Comparison of segmented and interpolated markers: *(a)* 1st frame, *(b)* 3rd frame, *(c)* 5th frame, *(d)* 7th frame

3.4 Conclusions

In this paper, a method of single stem cell extraction and monitoring in image plane in time sequence was proposed. The method is based on two-stage process of coarse segmentation of EF images followed by fine segmentation of BF ones. The coarse segmentation results in markers are applied in the fine segmentation step, which is based on the morphological watershed. Since EF images are acquired more rarely than BF images, only some of the markers can be computed directly. In order to create missing ones, the morphological interpolation combined with translation is used. This kind of interpolation allow morphing the shape and changing in position of the marker from the first given position into the next in a few intermediate points. The result of the proposed method applied to the experimental data captured as image sequences documenting the transfected neural stem cells' behavior in culture, was satisfied. Further investigations will focus on developing this method towards processing of sequences of higher cells density with multiple cells, also touching each other, as well as interfering and colliding during observations.

Acknowledgments

This work was supported by Ministry of Science and Higher Education of Poland, project no. 3911/B/T02/2008/34. We are grateful for neural stem cells line HUCB-NS from the NeuroRepair Dpt. Lab., MRI, PAS.

References

1. Iwanowski, M., Korzyńska, A.: Detection of the area covered by neural stem cells in cultures using textural segmentation and morphological watershed. In: ASC; Computer Recognition Systems, vol. 3, pp. 543–557. Springer, Heidelberg (2009)
2. Tse, S., Bradbury, L., Wan, J.W.L., Djambazian, H., Sladek, R., Hudson, T.: A combined watershed and level set method for segmentation of brightfield cell images. In: Proceedings of the SPIE, vol. 7258, pp. 72593G–72593G-10 (2009)
3. Iwanowski, M.: Metody morfologiczne w przetwarzaniu obraz"ow cyfrowych, AOW EXIT (2009)
4. Korzyńska, A., Iwanowski, M., Neuman, U., Dobrowolska, E., Hoser, P.: Comparison of the methods of microscopic image segmentation. In: Dossel, S. (ed.) WC 2009, IFMBE Proceedings, vol. 25/IV, pp. 425–428 (2009) ISBN 978-3-642-03897-6 (book), ISSN 1680-0737 (CD)
5. Selinummi, J., Ruusuvuori, P., Podolsky, I., Ozinsky, A., Gold, E., Yli-Harja, O., Aderem, A., Shmulevich, I.: Bright Field Microscopy as an Alternative to Whole Cell Fluorescence in Automated Analysis of Macrophage Images. PLoS ONE 4(10), e7497 (2009)
6. Ramoser, H.: Leukocyte segmentation and SVM classification in blood smear images. Machine Graphics and Vision 17(1/2), 187–200 (2008)
7. Piętka, B.D., Dulewicz, A., Jaszczak P.: Removing artefacts from microscopic Images of cytological smears. Machine Graphics and Vision 17(1/2),131–152 (2008)
8. Korzyńska, A., Zdunczuk, M.: Clustering as a method of image simplification. In: Pietka, K. (ed.) Inform. Tech. in Biomed., ASC, vol. 47, pp. 365–376 (2008)
9. Marciniak, A., Nieczkowski, T., Obuchowicz, A.: Color Homogram for segmentation of fine needle biopsy images. Machine Graphics and Vision 17(1/2), 153–165 (2008)
10. Witkowski, L.: A computer system for cells motility Evaluation. Machine Graphics and Vision 17(1/2), 167–186 (2008)
11. Markiewcz, T., Osowski, S.: Morphological operations for blood cells extraction from the image of the bone marrow smear. Przeglad Elektrotechniczny 84(5), 24–26 (2008)
12. Korzyńska, A., Iwanowski, M.: Detection of mitotic cell fraction in stem cells in cultures. In: ASC; Information Technologies in Biomedicine 1, vol. 47, pp. 365–376. Springer, Heidelberg (2008)
13. Korzyńska, A., Strojny, W., Hoppe, A., Wertheim, D., Hoser, P.: Segmentation of microscope images of living cells. Pattern Anal Applic. 10, 301–319 (2007)
14. Buzanska, L., Jurga, M., Stachowiak, E.K., Stachowiak, M.K., Domanska-Janik, K.: Neural Stem-like Cell Line Derived from a Nonhematopoietic Population of Human Umbilical Cord Blood. Stem Cell and Development 15, 391–406 (2006)
15. Boier Marti, I.M., Martineus, D.C.: Identification of spherical virus particles in digitized images of entire electron micrographs. Journal of Structural Biology 120, 146–157 (2005)

16. Koprowski, R., Wr"oblewski, Z.: Automatic segmentation of biological cell structures based on conditional opening and closing. Machine Graphics and Vision 14, 285–307 (2005)
17. Sabino, D.M.D., da F Costa, F., Costa, L., Rizzatti, E.G., Zago, M.A.: A texture approach to leukocyte recognition. Real-Time Imaging 10(4), 205–216 (2004)
18. Soille, P.: Morphological Image Analysis: Principles and Applications. Springer, Heidelberg (2004)
19. Jiang, K., Liao, Q.M., Dai, S.Y.: A novel white blood cell segmentation scheme using scale-space altering and watershed clustering. In: Proc. Int. Conf. on Machine Learning and Cybernetics, vol. 5, pp. 2820–2825 (2003)
20. Buzanska, L., Machaj, E.K., Zablocka, B.: Human Cord Blood - Derived Cells Attain Neuronal and Glial Features in Vitro. J. Cell Sci. 115, 2131–2138 (2002)
21. Periasamy, A.: Methods in Cellular Imaging. Oxford University Press, Oxford (2001)
22. Comaniciu, D., Meer, P.: Cell image segmentation for diagnostic pathology. In: Advanced Algorithmic Approaches to Medical Image Segmentation: State-of-the-Art Application in Cardiology, Neurology, Mammography and Pathology, pp. 541–558 (2001)
23. Iwanowski, M., Serra, J.: The morphological-affine object deformation. In: Mathematical Morphology and its Applications to Signal and Image Processing, pp. 81–90. Kluwer Academic Publishers, Dordrecht (2000)
24. Pham, D.L., Xu, C., Prince, J.L.: A survey of current methods in medical image segmentation, Vol. 2, pp. 315–338 (2000)
25. Serra, J.: Hausdorff distance and interpolations. In: Mathematical morphology and its Applications to Image and Signal Processing, pp. 107–114. Kluwer Academic Publishers, Dordrecht (1998)
26. Meyer, F.: A morphological interpolation method for mosaic images. In: Mathematical Morphology and its Applications to Image and Signal Processing, pp. 337–344. Kluwer Academic Publishers, Dordrecht (1996)
27. Serra, J.: Image analysis and mathematical morphology, vol. 1. Academic Press, London (1983)
28. Fu, K.S., Mui, J.K.: A survey on image segmentation. Pattern Recognition 13(1), 3–16 (1981)

Shape Representation and Shape Coefficients via Method of Hurwitz-Radon Matrices

Dariusz Jakóbczak

Department of Electronics and Computer Science
Technical University of Koszalin, Śniadeckich 2, 75-453 Koszalin, Poland
djakob@ie.tu.koszalin.pl

Abstract. Computer vision needs suitable methods of shape representation and contour reconstruction. One of them called method of Hurwitz-Radon Matrices (MHR) can be used in representation and reconstruction of shapes of the objects in the plane. Another problem is connected with shape coefficients. This paper contains the way of length estimation and area estimation via MHR method. Proposed method is based on a family of Hurwitz-Radon (HR) matrices. The matrices are skew-symmetric and possess columns composed of orthogonal vectors. The operator of Hurwitz-Radon (OHR), built from these matrices, is described. The shape is represented by the set of nodes. It is shown how to create the orthogonal and discrete OHR and how to use it in a process of shape representation and reconstruction. MHR method is interpolating the curve point by point without using any formula or function.

Keywords: Shape representation, curve interpolation, contour reconstruction, shape coefficients, length estimation, area estimation, Hurwitz-Radon matrices.

1 Introduction

A significant problem in machine vision and computer vision [1] is that of appropriate shape representation and reconstruction. Classical discussion about shape representation is based on the problem: contour versus skeleton. This paper is voting for contour which forms boundary of the object. Contour of the object, represented by contour points, consists of information which allows us to describe many important features of the object as the shape coefficients [2].

Digital curve (open or closed) may be represented by chain code (Freeman's code). Chain code depends on selection of the started point and transformations of the object. So Freeman's code is one of the method how to describe and to find contour of the object. An analog (continuous) version of Freeman's code is the curve α - s. Another contour representation and reconstruction is based on the Fourier coefficients calculated in Discrete Fourier Transformation (DFT). These coefficients are used to fix similarity of the contours with different sizes or directions. If we assume that contour is built from segments of a line and fragments of circles or ellipses, Hough transformation is applied to detect the

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 411–419, 2010.

contour lines. Also geometrical moments of the object are used during the process of object shape representation [3]. MHR method requires to detect specific points of the object contour, for example in compression and reconstruction of monochromatic medical images [4]. Contour is also applied in the shape decomposition [5]. Many branches of medicine, for example computed tomography [6], need suitable and accurate methods of contour reconstruction [7]. Also industry and manufacturing are looking for the methods connected with geometry of the contour [8]. So suitable shape representation and precise reconstruction or interpolation [9] of the object contour is a key factor in many applications of computer vision.

2 Shape Representation

The shape can be represented by the object contour, i.e. curves that create each part of the contour. One curve is described by the set of nodes $(x_i, y_i) \in \mathbf{R}^2$ (characteristic points) as follows in proposed method:

1. Nodes (interpolation points) are settled at local extrema (maximum or minimum) of one of coordinates and at least one point between two successive local extrema;
2. Nodes (xi,yi) are monotonic in coordinates xi ($xi < xi$+1 for all i) or yi ($yi < yi$+1);
3. One curve (one part of the contour) is represented by at least five nodes.

Condition 1 is done for the most appropriate description of a curve. So we have n curves $C1$, $C2$, ... Cn that build whole contour and each curve is represented by the nodes according to assumptions 1-3.

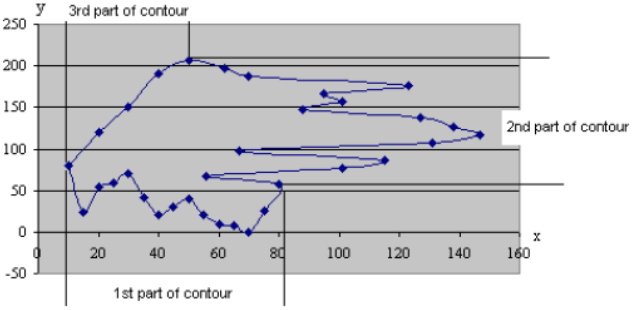


Fig. 1. A contour consists of three parts (three curves and their nodes)

Fig. 1 is an example for $n = 3$: first part of the contour $C1$ is represented by the nodes monotonic in coordinates xi, second part of the contour $C2$ is represented by the nodes monotonic in coordinates yi and third part $C3$ could be represented by the nodes either monotonic in coordinates xi or monotonic in coordinates *yi*. Number of the curves is optional and number of the nodes for each curve is optional too (but at least five nodes for one curve). Representation points are treated as interpolation nodes. How accurate can we reconstruct whole contour using representation points? The shape reconstruction is possible using novel MHR method.

3 Shape Reconstruction

The following question is important in mathematics and computer sciences: is it possible to find a method of curve interpolation in the plane without building the interpolation polynomials or other functions? Our paper aims at giving the positive answer to this question. In comparison MHR method with Bézier curves, Hermite curves and B-curves (*B-splines*) or NURBS one unpleasant feature of these curves must be mentioned: small change of one characteristic point can make big change of whole reconstructed curve. Such a feature does not appear in MHR method. The methods of curve interpolation based on classical polynomial interpolation: Newton, Lagrange or Hermite polynomials and the spline curves which are piecewise polynomials [10]. Classical methods are useless to interpolate the function that fails to be differentiable at one point, for example the absolute value function $f(x) = x$at x=0. If point (0;0) is one of the interpolation nodes, then precise polynomial interpolation of the absolute value function is impossible. Also when the graph of interpolated function differs from the shape of polynomials considerably, for example $f(x) = 1/x$, interpolation is very hard because of existing local extrema of polynomial. Lagrange interpolation polynomial for function $f(x) = 1/x$ and nodes (5;0.2), (5/3;0.6), (1;1), (5/7;1.4), (5/9;1.8) has one minimum and two roots.

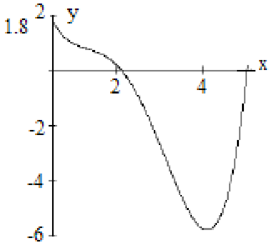


Fig. 2. Lagrange interpolation polynomial for nodes (5;0.2), (5/3;0.6), (1;1), (5/7;1.4), (5/9;1.8) differs extremely from the shape of function $f(x) = 1/x$

We cannot forget about the Runge's phenomenon: when the interpolation nodes are equidistance then high-order polynomial oscillates toward the end of the interval, for example close to -1 and 1 with function $f(x) = 1/(1+25x2)$ [11]. Method of Hurwitz – Radon Matrices (MHR), described in this paper, is free of these bad examples. The curve or function in MHR method is parameterized for value $\alpha \in [0;1]$ in the range of two successive interpolation nodes.

3.1 The Operator of Hurwitz-Radon

Adolf Hurwitz (1859-1919) and Johann Radon (1887-1956) published the papers about specific class of matrices in 1923, working on the problem of quadratic forms. Matrices Ai**,** i = 1,2… m satisfying

$$A_jA_k + A_kA_j = 0, A_j^2 = -I\mathrm{for} j \neq k; j,k = 1,2...m$$

are called *a family of Hurwitz - Radon matrices*. A family of Hurwitz - Radon (HR) matrices has important features [12]: HR matrices are skew-symmetric (AiT = - Ai) and reverse matrices are easy to find (Ai-1 = - Ai). Only for dimension N = 2, 4 or 8 the family of HR matrices consists of N - 1 matrices. For N = 2 we have one matrix:

$$A_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

For N = 4 there are three HR matrices with integer entries:

$$A_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, A_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, A_3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}.$$

For N = 8 we have seven HR matrices with elements 0, ±1.

So far HR matrices are applied in electronics [13]: in Space-Time Block Coding (STBC) and orthogonal design [14], also in signal processing [15] and Hamiltonian Neural Nets [16].

If one curve is described by a set of representation points $\{(xi,yi), i = 1, 2, \ldots, n\}$ monotonic in coordinates xi, then HR matrices combined with the identity matrix IN are used to build the orthogonal and discrete Hurwitz - Radon Operator (OHR). For nodes ($x1$,$y1$), ($x2$,$y2$) OHR M of dimension N = 2 is constructed:

$$B = (x_1 \cdot I_2 + x_2 \cdot A_1)(y_1 \cdot I_2 - y_2 \cdot A_1) = \begin{bmatrix} x_1 & x_2 \\ -x_2 & x_1 \end{bmatrix} \begin{bmatrix} y_1 & -y_2 \\ y_2 & y_1 \end{bmatrix}, M = \frac{1}{x_1^2 + x_2^2} B,$$

$$M = \frac{1}{x_1^2 + x_2^2} \begin{bmatrix} x_1y_1 + x_2y_2 & x_2y_1 - x_1y_2 \\ x_1y_2 - x_2y_1 & x_1y_1 + x_2y_2 \end{bmatrix}. \tag{1}$$

Matrix M in (1) is found as a solution of equation:

$$\begin{bmatrix} a & b \\ -b & a \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}. \tag{2}$$

For nodes ($x1$,$y1$), ($x2$,$y2$), ($x3$,$y3$), ($x4$,$y4$), monotonic in xi, OHR of dimension N = 4 is constructed:

$$M = \frac{1}{x_1^2 + x_2^2 + x_3^2 + x_4^2} \begin{bmatrix} u_0 & u_1 & u_2 & u_3 \\ -u_1 & u_0 & -u_3 & u_2 \\ -u_2 & u_3 & u_0 & -u_1 \\ -u_3 & -u_2 & u_1 & u_0 \end{bmatrix} \tag{3}$$

where

$$u_0 = x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4, u_1 = -x_1y_2 + x_2y_1 + x_3y_4 - x_4y_3,$$
$$u_2 = -x_1y_3 - x_2y_4 + x_3y_1 + x_4y_2, u_3 = -x_1y_4 + x_2y_3 - x_3y_2 + x_4y_1.$$

Matrix M in (3) is found as a solution of equation:

$$\begin{bmatrix} a & b & c & d \\ -b & a & -d & c \\ -c & d & a & -b \\ -d & -c & b & a \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}. \tag{4}$$

For nodes ($x1$,$y1$), ($x2$,$y2$), ..., ($x8$,$y8$), monotonic in xi, OHR of dimension $N = 8$ is built [17] similarly as (1) or (3). Note that OHR operators M (1)-(3) satisfy the condition of interpolation

$$M \cdot \mathbf{x} = \mathbf{y} \tag{5}$$

for $\mathbf{x}$=($x1$,$x2$..., xN) T∈ $\boldsymbol{RN}$,$\mathbf{x}\neq$ $\mathbf{0}$,$\mathbf{y}$=($y1$,$y2$..., yN) T∈$\boldsymbol{RN}$,N=2,4 or 8.

If one curve is described by a set of nodes $\{(xi,yi),\ i = 1, 2, \ldots, n\}$ monotonic in coordinates yi, then HR matrices combined with the identity matrix IN are used to build the orthogonal and discrete reverse Hurwitz - Radon Operator (reverse OHR)M-1. If matrix M in (1)-(3) is described as:

$$M = \frac{1}{\sum_{i=1}^{N} x_i^2}(u_0 \cdot I_N + D),$$

where D with elements $u1$, ..., uN-1, then reverse OHR M-1 is given by:

$$M^{-1} = \frac{1}{\sum_{i=1}^{N} y_i^2}(u_0 \cdot I_N - D). \tag{6}$$

Note that reverse OHR operator (6) satisfies the condition of interpolation

$$M - 1 \cdot \mathbf{y} = \mathbf{x}\ \mathbf{(7)} \tag{7}$$

for $\mathbf{x}$=($x1$,$x2$...,xN) T∈ $\boldsymbol{RN}$, $\mathbf{y}$=($y1$,$y2$...,yN) T∈ $\boldsymbol{RN}$, $\mathbf{y}\neq$ $\mathbf{0}$, N =2,4 or 8.

3.2 MHR Method

Key question looks as follows: how can we compute coordinates of points settled between the interpolation nodes? On a segment of a line every number "c" situated between "a" and "b" is described by a linear (convex) combination c=α . a+(1 - α) . b for

$$\alpha = \frac{b-c}{b-a} \in [0;1]. \tag{8}$$

When the nodes are monotonic in coordinates xi, the average OHR operator $M2$ of dimension $N = 2$, 4 or 8 is constructed as follows:

$$M_2 = \alpha \cdot M_0 + (1 - \alpha) \cdot M_1 \tag{9}$$

with the operator $M0$ built (1)-(3) by "odd" nodes ($x1$=a,$y1$), ($x3$,$y3$), ..., ($x2N$-1,$y2N$-1) and $M1$ built (1)-(3) by "even" nodes ($x2$=b,$y2$), ($x4$,$y4$), ..., ($x2N$,$y2N$). Having the operator $M2$ for coordinates $xi < xi{+}1$ it is possible to reconstruct the second coordinates of points (x,y) in terms of the vector C defined with

$ci = \alpha.x2i$-1+ (1-α).$x2i$,$i = 1, 2,\ldots, N$ **(10)**

as $C = [c1,\ c2,\ldots,\ cN]$T. The required formula is similar to (5):

$$Y(C) = M_2 \cdot C \tag{10}$$

in which components of vector $Y(C)$ give the second coordinate of the points (x,y) corresponding to the first coordinate, given in terms of components of the vector C.

On the other hand, having the operator $M2$-1 for coordinates $yi < yi{+}1$ it is possible to reconstruct the first coordinates of points (x,y):

$M_2^{-1} = \alpha \cdot M_0^{-1} + (1-\alpha) \cdot M_1^{-1}$, $ci = \alpha.y2i$-1+ (1-α).$y2i$,

$$X(C) = M_2^{-1} \cdot C. \tag{11}$$

Contour of the object is constructed with several number of curves. Calculation of unknown coordinates for contour points using (8)-(11) is called by author the method of Hurwitz - Radon Matrices (MHR). Here is the application of MHR method for functions $f(x) = 1/x$ (nodes as Fig. 2) and $f(x) = 1/(1{+}25x2)$ with five nodes equidistance in first coordinate: xi = -1, -0.5, 0, 0.5, 1.

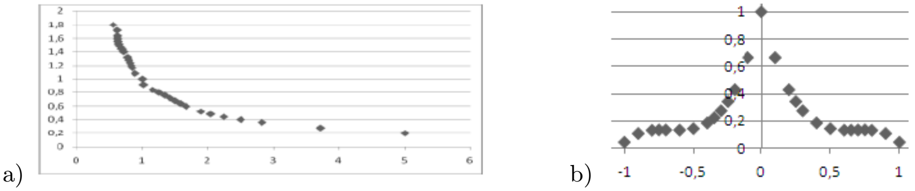


Fig. 3. Twenty six interpolated points of functions $f(x)$=1/x (a) and $f(x)$ = 1/(1+25x2) (b) using MHR method with 5 nodes

MHR interpolation for function $f(x) = 1/x$ gives better result then Lagrange interpolation (Fig. 2). The same can be said for function $f(x) = 1/(1{+}25x2)$.

4 Shape Coefficients

Some of the shape coefficients are calculated using area of the object S and length of the contour L. For example:

$$R_S = \frac{L^2}{4\pi S}, R_{C1} = 2\sqrt{\frac{S}{\pi}}, R_{C2} = \frac{L}{\pi}, R_M = \frac{L}{2\sqrt{\pi S}} - 1.$$

Area S is also used in coefficients of: Blair-Bliss, Danielsson, compactness [18].

The contour is divided into n curves $C1$, $C2$, ... Cn. Having nodes ($x1$,$y1$), ($x2$,$y2$),…, (xm,ym) for each Ci in MHR method, it is possible to compute as many curve points as we want for any parameter $\alpha \in [0;1]$. Assume that k is the number of reconstructed points together with m nodes. So a curve Ci consists of k points that could be indexed ($x1$',$y1$'), ($x2$',$y2$'),…, (xk',yk'), where ($x1$',$y1$') = ($x1$,$y1$) and (xk',yk') = (xm,ym).The length of a curve Ci, consists of k points, is estimated:

$$d(C_i) = \sum_{i=1}^{k-1} \sqrt{(x'_{i+1} - x'_i)^2 + (y'_{i+1} - y'_i)^2}. \tag{12}$$

Length of whole contour L is computed:

$L = d(C1) + d(C2) + \ldots + d(Cn)$. **(14)**

Area of the object can be divided horizontally or vertically into the set of l polygons: triangles, squares, rectangles, trapezoids.

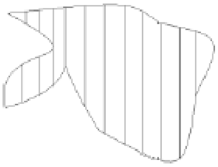


Fig. 4. The object area consists of polygons

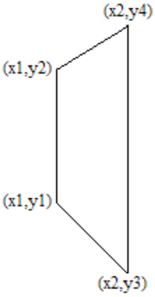


Fig. 5. Trapezoid as a part of the object

The coordinates of corners for each polygon Pi are calculated by MHR method and then it is easy to estimate the area of Pi. For example $P1$ as a trapezoid with the corners ($x1$,$y1$), ($x1$,$y2$), ($x2$,$y3$), ($x2$,$y4$):

Area of a trapezoid $P1$ is computed:

$$s(P_1) = \frac{1}{2} |x_2 - x_1| \cdot (|y_2 - y_1| + |y_4 - y_3|). \tag{13}$$

Estimation of the object area S is given by a formula:

$$S = \sum_{i=1}^{l} s(P_i). \tag{14}$$

Contour points, calculated by MHR method [19], are applied in shape coefficients.

5 Conclusions

The method of Hurwitz-Radon Matrices leads to contour interpolation and shape reconstruction depending on the number and location of representation points. No characteristic features of curve are important in MHR method: failing to be differentiable at any point, the Runge's phenomenon or differences from the shape of polynomials. These features are very significant for classical polynomial interpolations. MHR method gives the possibility of reconstruction a curve consists of several parts, for example closed curve (contour). The only condition is to have a set of nodes for each part of a curve or contour according to assumptions in MHR method. Shape representation and curve reconstruction by MHR method is connected with possibility of changing the nodes coordinates and reconstruction of new curve or contour for new set of nodes, no matter what shape of curve or contour is to be reconstructed. Main features of MHR method are: accuracy of shape reconstruction depending on number of nodes and method of choosing nodes; reconstruction of curve consists of L points is connected with the computational cost of rank $O(L)$;

MHR method is dealing with local operators: average OHR operators are built by successive 4, 8 or 16 nodes, what is connected with smaller computational costs then using all nodes; MHR is not an affine interpolation [19].

Future works are connected with: geometrical transformations of contour (translations, rotations, scaling)- only nodes are transformed and new curve (for example contour of the object) for new nodes is reconstructed, possibility to apply MHR method to three-dimensional curves and connection MHR method with object recognition.

References

1. Ballard, D.H.: Computer Vision. Prentice Hall, New York (1982)
2. Tadeusiewicz, R., Flasiński, M.: Image Recognition. PWN, Warsaw (1991)
3. Chora"s, R.S.: Computer Vision. Exit, Warsaw (2005)
4. Jakóbczak, D., Kosiński, W.: Application of Hurwitz - Radon Matrices in Monochromatic Medical Images Decompression. In: Kowalczuk, Z., Wiszniewski, B. (eds.) Intelligent Data Mining in Diagnostic Purposes: Automatics and Informatics, pp. 389–398. PWNT, Gdansk (2007)
5. Latecki, L.J., Lakaemper, R.: Convexity Rule for Shape Decomposition Based on Discrete Contour Evolution. Computer Vision and Image Understanding 3(73), 441–454 (1999)
6. Cierniak, R.: Computed Tomography. Exit, Warsaw (2005)
7. Soussen, C., Mohammad-Djafari, A.: Polygonal and Polyhedral Contour Reconstruction in Computed Tomography. IEEE Transactions on Image Processing 11(13), 1507–1523 (2004)
8. Tang, K.: Geometric Optimization Algorithms in Manufacturing. Computer – Aided Design & Applications 2(6), 747–757 (2005)
9. Kozera, R.: Curve Modeling via Interpolation Based on Multidimensional Reduced Data. Silesian University of Technology Press, Gliwice (2004)
10. Dahlquist, G., Bjoerck, A.: Numerical Methods. Prentice Hall, New York (1974)

11. Ralston, A.: A First Course in Numerical Analysis. McGraw-Hill Book Company, New York (1965)
12. Eckmann, B.: Topology, Algebra, Analysis- Relations and Missing Links. Notices of the American Mathematical Society 5(46), 520–527 (1999)
13. Citko, W., Jakóbczak, D., Sieńko, W.: On Hurwitz - Radon Matrices Based Signal Processing. In: Workshop Signal Processing at Poznan University of Technology (2005)
14. Tarokh, V., Jafarkhani, H., Calderbank, R.: Space-Time Block Codes from Orthogonal Designs. IEEE Transactions on Information Theory 5(45), 1456–1467 (1999)
15. Sieńko, W., Citko, W., Wilamowski, B.: Hamiltonian Neural Nets as a Universal Signal Processor. In: 28th Annual Conference of the IEEE Industrial Electronics Society IECON (2002)
16. Sieńko, W., Citko, W.: Hamiltonian Neural Net Based Signal Processing. In: The International Conference on Signal and Electronic System ICSES (2002)
17. Jakóbczak, D.: 2D and 3D Image Modeling Using Hurwitz-Radon Matrices. Polish Journal of Environmental Studies 4A(16), 104–107 (2007)
18. Tadeusiewicz, R., Korohoda, P.: Computer Analysis and Image Processing. FPT, Cracow (1997)
19. Jakóbczak, D.: Curve Interpolation Using Hurwitz-Radon Matrices. Polish Journal of Environmental Studies 3B(18), 126–130 (2009)

Characteristics of Architectural Distortions in Mammograms - Extraction of Texture Orientation with Gabor Filters

Magdalena Jasionowska, Artur Przelaskowski, and Rafał Jóźwiak

Institute of Radioelectronics, Warsaw University of Technology
Nowowiejska 15/19, Warsaw, Poland
mjasiono@ire.pw.edu.pl, arturp@ire.pw.edu.p
http://www.ire.pw.edu.pl/~arturp/

Abstract. Differentiation of pathological architectural distortions (ADs) in mammograms is investigated with the aid of numerical description of local texture orientation. Challenging purpose of our long-term research is effective method of ADs detection in mammography. It is reasoned by significant limitations in efficiency of commercial and research detectors of pathological ADs. Extracted and enhanced texture structures are analysed in order to automatically indicate the small regions with potential pathologies. Gabor filtering is proposed as promising extractor of texture orientation according to our previous studies and the reported results in literature. However, adjusting Gabor filters (GF) to extremely changing manifestation of ADs spicules, diversified in size, forms, shapes and intensity is still open question. Therefore, we optimize GF by impulse response shaping, adjusting of angular resolution and taking into account local image activity. Finally, a *probability map* of ADs appearance is estimated using the results of Gabor filtering of mammograms. Experimental verification of Gabor maps efficiency leads to 82% sensitivity at 2.64 false positives per test case (FPR) if we use only Gabor filtering to detect ADs or 88% sensitivity at $FPR = 4.4$ in case of additionally ADs recognition stage on selected ROIs by texture orientation enhancement, analysis and extraction.

1 Introduction

Breast cancer survival depends on its earliest possible detection. The goal of screening mammography is to find very subtle pathology signs which appear in the early stadium of breast disease. One of the typical types of abnormal findings is architectural distortion (AD)[5]. It is a breast lesion in which the normal structure of the breast parenchyma is distorted as if being pulled into a central point, without a visible central density (Fig.1) [2].

For physicians, it is more difficult to detect ADs than any other findings (i.e. microcalcifications or masses) because of their low subtlety as well as diversified and indefinite manifestation. Moreover, ADs are difficult in specific interpretation. In screening mammography only 3% of lesions detected by radiologists are

L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 420–430, 2010.

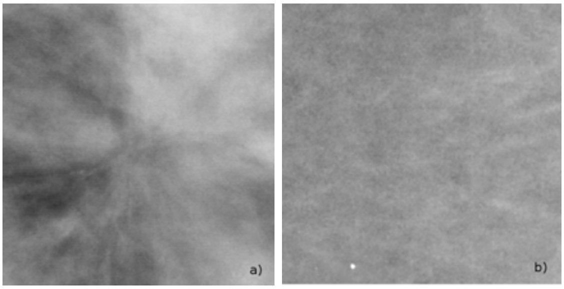


Fig. 1. Two ROIs of original mammograms: (a) with AD - abnormal finding with spiculations radiating and (b) with normal breast tissue [4]

ADs, whereas 61% of detected findings are masses and 90% of all lesions are microcalcifications [3]. Commercial computer - aided detection (CAD) systems are able to detect certain types of architectural distortions, but their sensitivity is not satisfactory (the sensitivity of Image Chacker R2 Technology equals 49% by number of false positives per image $FPI = 0.7$ [9]). There is a problem with indefinite characteristic features of ADs. In image processing, this abnormal finding is generally modeled as a group of line structures of different directionality. Bovik et al. [12] used filtering in the Radon domain as the extractor of texture orientation. To enhance linear structures (spicules of ADs and spiculated masses) the Radon transform of the mammogram was applied with following filtering in the Radon domain. The basic idea was to design filters matched to pattern findings. Endo et al. [6] detected the suspect regions with ADs by concentration estimated indexes of line structures. The indexes were extracted using mean curvature - areas with ADs corresponded to the high concentration regions. Ayres et al. [1,10,11] filtered mammograms with a bank of Gabor filters of different texture orientation angles. Resulting ADs detection performance was significantly high. Similarly, according to the study of Ma et al. [8] the average rate of image texture extraction based on Gabor filtering was satisfactory - approximately 74%.

After initial experiments, Gabor filtering (GF) was concluded as promising processing procedure to enhance texture orientation. Consequently, we designed ADs detection method called ArDist [7] that was based on histogram analysis of Gabor probability maps resulting in indications of regions of suspected pathologies. The key problem was design of Gabor filters adjusted to representative and invariant ADs patterns with high enough sensitivity. Automatically indicated small regions of high ADs susceptibility or probability were the results of optimized GF-based method (ADGF).

1.1 ADs Detection Method

The assumptions, general concept and implementation of ADs detection method was presented in [7]. A schematic diagram of ArDist method was given in Fig. 2.

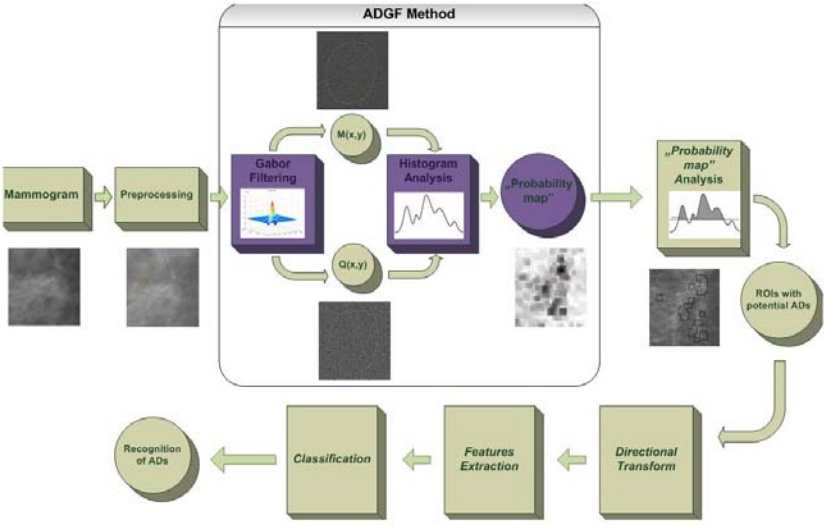


Fig. 2. Schema of ArDist method that consists of three common stages: preprocessing, ADGF method based on Gabor filtering of mammograms and the recognition of ADs in selected ROIs. Selected blocks optimization was the goal of presented research.

It consists of three common stage: i) preprocessing of mammogram, ii) detecting of ROIs with potential pathologies (ADGF method), iii) recognition of ADs in selected ROIs.

The preprocessed mammogram (after reducing the noise and the influence of the low-frequency components of the mammogram) was filtered with 9 × 9 Gabor filters bank of 180 different orientations. The Gabor filter - a sinusoidally modulated Gaussian functions is defined as follows:

$$g_k(x,y) = \exp([-\frac{1}{2}[(\frac{x'}{S'_x})^2 + (\frac{y'}{S'_y})^2]]) * \cos(2\pi f x') \tag{1}$$

with $x' = x\cos\theta_k + y\sin\theta_k$, $y' = y\cos\theta_k - x\sin\theta_k$, $\theta_k \in [-\pi/2, \pi/2]$ donates the angle of texture orientation, $f = 0.075$ indicates the frequency of sinusoidal function and $S_x = 4$, $S_y = 4$ parameters determine width and elongation of Gabor filter respectively (Fig.3a).

As a result we received two Gabor maps - amplitude and angular maps. The map with the maximum values of amplitude $M(x,y)$ was estimated as the highest magnitude of amplitude among 180 values of different orientations (for 180 corresponding filter responses). Angular map contains the dominant angles of texture orientation $Q(x,y)$, i.e. the angles for which the magnitude of amplitudes is maximum (Fig.4b and Fig.4c respectively).

Next, in order to select ROIs with potential ADs, block-based histogram analysis of Gabor maps (with 100×100 block size and the shift step equals 50 pixels) and creating the *probability map* are employed. The *probability map* (Fig.5b) is

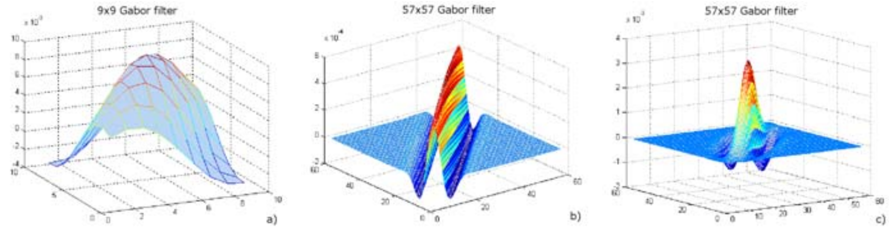


Fig. 3. Gabor filter mask with $\theta_k = \pi/2$ and various parameters: (a) 9x9 mask, $S_x = 4$, $S_y = 4$, $f = 0.075$, (b) 57x57 mask, $S_x = 4$, $S_y = 80$, $f = 0.075$, (c) 57x57 mask, $S_x = 4$, $S_y = 12$, $f = 0.075$

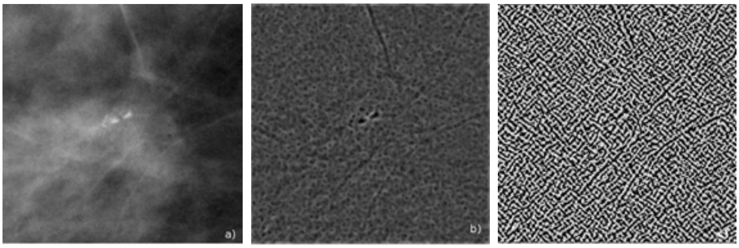


Fig. 4. Gabor maps - the images of extracted texture orientation: (a) ROI of original mammogram, (b) ROI of the maximal amplitude $M(x, y)$, (c) ROI of the dominant angles of texture orientation $Q(x, y)$

created by assignment the highest probability of ADs appearance (reflected as a maximum gray scale value) to blocks, in which after the histogram analysis the number n of dominant texture orientation angles is equal or higher than the number $N = 17$ (set experimentally) which characterizes ADs multidirectionality. Next, the blocks (classified by simple thresholding) with the high probability of ADs occurrence are marked with stars (Fig.5c).

The results of ADGF method - selected ROIs of diverse local directions were used to detect probable ADs according to the following paradigm: i) more sensitive directional region transformation based on the 2D Fourier transform in polar coordinates to enhance only strong spicules, ii) extraction and selection of specific features, iii) adjusted features classification (Fig. 2). To differentiate the ADs from normal breast tissue two groups of features were used: the image-domain texture features and angular local features computed in directional transform domain. The support vector classifier with linear kernel and regularization was used for final classification.

Efficiency of ADs Detection. We found proposed method quite effective in comparison to the commercial CAD systems. The efficiency of ArDist method was 68% of sensitivity at the $FPI = 0.86$. Especially, the sensitivity of ADGF stage was 79% at high rate of $FPI = 7.8$. In order to increase ADs detection

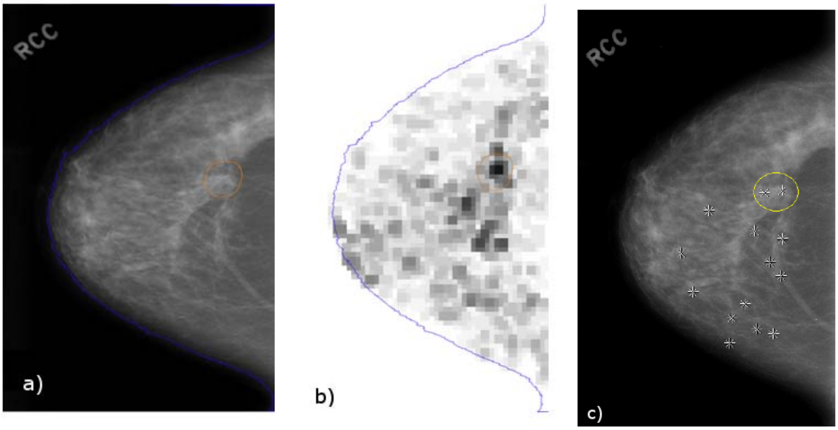


Fig. 5. The result of ADGF method: (a) the original mammogram with abnormal finding, (b) the created *probability map*, (c) the original mammograms with marked by stars blocks of the high probability of ADs occurrence. The circle indicates the sites of AD, marked by radiologist.

efficiency, sensitivity of key ADGF stage should be increased whereas the number of false positives is expected to be significantly reduced. Reported research was concentrated on the analysis of Gabor filters with adjusting impulse response parameters and followed histogram-based data analysis to enhance ADs spicules and increase selectivity of resulting *probability maps.*

2 Optimization of ADGF Method

The considered ADGF method consists of two fundamental steps which were optimized: extraction of ADs texture orientation with Gabor filtering and directional imaged tissue characteristics by local histogram-based analysis. Resulting *probability map* is used to select the regions of high probability of ADs. The process of ADGF optimization is outlined in the following subsections.

2.1 Gabor Filtering

According to previous experiments, Gabor filtering was proved to be very sensitive method of texture orientation description. However, more selective directional tissue characteristic is necessary to distinguish pathological distortions from areas of normal breast tissue. In ADs detection, it is very important to pass over all fine structures (considered as a noise) and to enhance only the long narrow ADs spicules radiating in different directions of small (local) regions. Impulse responses of designed Gabor filters should be flexibly adjusted to the approximated models of AD spicules. The filter parameters should be matched to provide enough resolution so that two neighboring spicules are distinguishable,

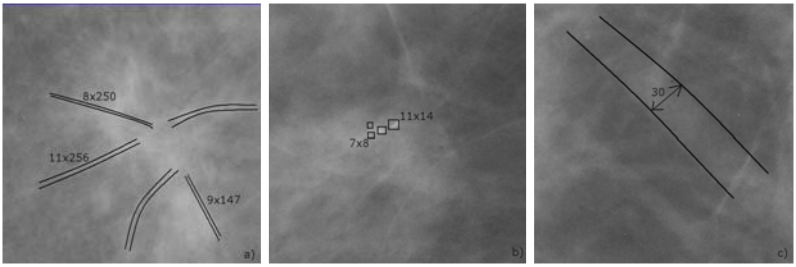


Fig. 6. ROI with some marked structures on mammogram: the spicules of ADs (a) and false, natural elements in detecting process of ADs - microcalcifications (b) and lactic ducts (c). The size of individual structures is given in pixels.

but the edges of the same spicule with changing width are not treated as two separate line structures. Concurrently, the other less outlined and shorter spicules of different direction should be masked. To optimize GF and consequently to assure high AD selectivity we need representative AD models. It is estimated for test set of fixed resolution mammograms that the width of spicule can be 9 pixels in average and the average minimal length can be 80 pixels (Fig.6). Such model is scalable considering resolution of imaging systems.

According to chosen definition of Gabor filters (i.e. general formula given in Eq. 1) and presented assumptions and requirements, the best filter parameters were determined as follows:

- parameter determining width of Gabor filter $S_x = 4$ which contributes to enhancing only the narrow spicules and eliminating the effect of detecting the edges of the same spicule as two separate line structures;
- elongation in the y direction $S_y = 80$ in order to extract long enough spicules of ADs and eliminate all fine structures; to get a such elongation the filter mask is fixed to 57×57 pixel block;
- the frequency of sinusoidal function $f = 0.075$ which causes no additional spines being result of sinusoidally modulated Gaussian functions (Fig.3b).

To verify the results of such Gabor filtering, simulated phantoms of ADs and false natural structures were considered. Exemplary results of simple experiments were presented in Fig.7.

However, the extraction of local dominant directions in real mammograms is too distinct, causing the effect masking of other, less outlined directions - see the example of masking effects in Fig. 8. Therefore, we modified parameters of Gabor filters. Modified AD model includes context dependencies of crossing or closely located spicules. On the basis of experiments, we accepted the parameter that determine the elongation in y direction $S_y = 12$ (Fig. 3c). Then, we get rid of the negative effect of less outlined direction masking. The results of mammogram filtering with modified GF parameters were shown in Fig. 9. Because of high computational complexity of Gabor filtering with mask of 57×57 pixels and possible reduction of angular resolution, we limited the

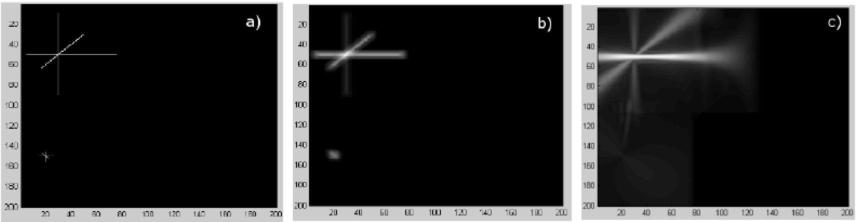


Fig. 7. The example of optimized GF verification: (a) phantom of AD (crossing of longer elements) and false natural structure (crossing of short elements); the result of Gabor filtering with selected parameters: (b) initially used filter mask of 9 × 9 pixels with $S_x = 4$, $S_y = 4$, $f = 0.075$, (c) adjusted to AD model filter mask of 57 × 57 pixels with $S_x = 4$, $S_y = 80$, $f = 0.075$. Good result (only enhanced AD and skipped false element) is achieved in (c).

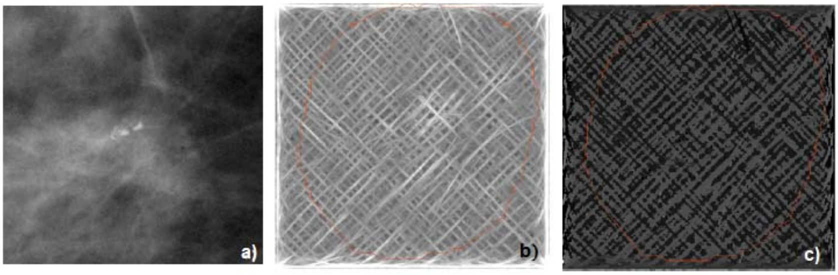


Fig. 8. The effect of distinct extraction only dominant directions caused by Gabor filtering with the following parameters: mask of 57 × 57 pixels, $S_x = 4$, $S_y = 80$, $f = 0.075$: (a) ROI of original mammogram, (b) $M(x, y)$ map, (c) $Q(x, y)$ map.

number of the texture orientations to 60 angles from range of $[-\pi/2 : \pi/2]$, i.e. $\theta_k \in [-\pi/2 : \pi/60 : \pi/2]$. Typically, single spicules cover a few neighbor directions of angular analysis in full $[-\pi/2 : \pi/2]$ range. Thus, during histogram-based analysis of Gabor maps one spicule is represented in histogram as a few near-lying peaks, which are labeled with one direction valued in angle average. Therefore, we can limit the number of texture orientation angles without risk of loss of some spicule direction in mammogram, what is experimentally verified and confirmed (Fig. 10). Finally, Gabor filters with the following parameters were used as the most effective for AD selective representation: mask of 57 × 57 pixels, $S_x = 4$, $S_y = 12$, $f = 0.075$ and $\theta_k \in [-\pi/2 : \pi/60 : \pi/2]$.

2.2 Histogram Analysis

The Gabor maps of $M(x, y)$ and $Q(x, y)$ were analyzed based on local histogram estimates and adaptive selection to indicate regions of high probability of ADs appearance. Final effect was *probability map* used for indication of regions with

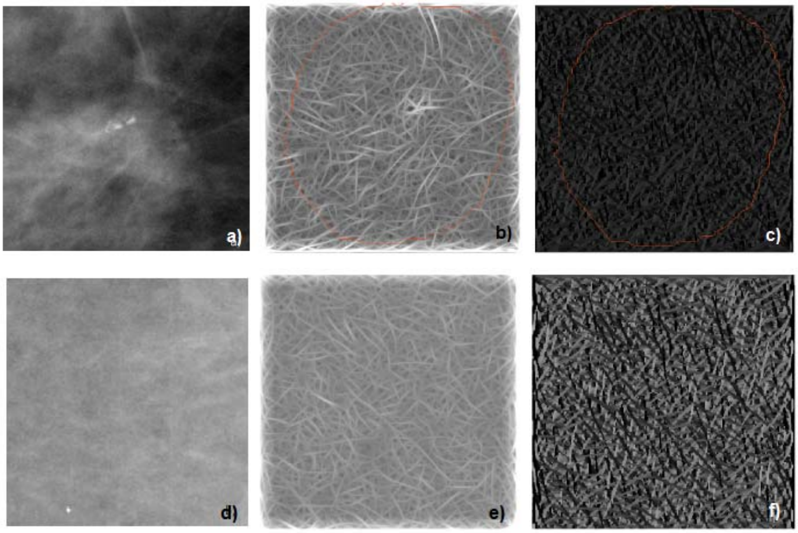


Fig. 9. The result of Gabor filter with the following parameters: mask of 57×57 pixels, $S_x = 4$, $S_y = 12$, $f = 0.075$ for region: (a) with abnormal finding and (d) with normal breast tissue. $M(x, y)$ maps are presented respectively in (b) and (e), $Q(x, y)$ maps in (c) and (f). Differentiation of Gabor maps for these regions is even visually convincing.

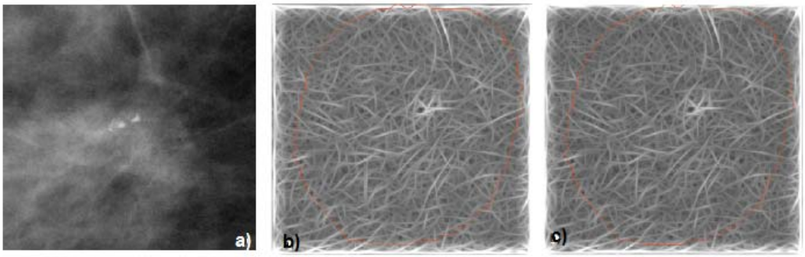


Fig. 10. The comparison of the effect of selected region filtering with different resolution: (a) ROI of original mammogram, (b) ROI after Gabor filtering of 60 angles of texture orientations and (c) ROI after Gabor filtering of 180 angles of texture orientations

potential ADs (Fig.2). The Gabor maps were analyzed by regions of $K \times K$ pixels, where:

a) the number of **context dominant angles** of texture orientation is computed in overlapping blocks B_k of $k \times k$ pixels, where $k = K/5$; the dominant texture orientation angles are defined as a set of angles $Q(x, y) \in B_k$ for which the amplitude values of $M(x, y)$ are higher than $t_M = 0.8 * \max_{(x,y)\in B_k} M(x, y)$;
b) only texture orientation angles with high occurrence number across the whole region of $K \times K$ pixels are selected for *probability map* creation; the **region**

dominant angles of $K \times K$ block histogram H_K are estimated by thresholding with $t_H = 500$;

c) the *probability map* is created in two ways:
   - option 1: the number of **region dominant angles** for each block $K \times K$ is scaled to gray scale $0 - 255$ (Fig.11b);
   - option 2: the blocks with the number of **region dominant angles** higher than characteristic for ADs number $N = 25$ (experimentally established in our preliminary analysis on ROIs with ADs, selected from "gold-standard"cases) take the maximum gray scale value and for the other blocks the scale is extended as follows: $(n * MAX_{scale})/MAX_{scale}$ (Fig.12b);

d) analyzing of the *probability map* to enhance only blocks the high probability of ADs appearance (Fig.11c,12c), consequently the ROIs with potential pathologies.

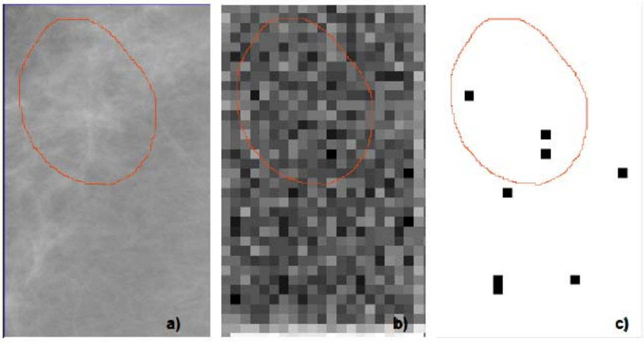


Fig. 11. The ROI of original mammogram (a) and the *probability map* established by option 1 (b) - it is displayed in inverse scale. The thresholded *probability map* with the black blocks which are characterized by the highest probability of ADs appearance (c).

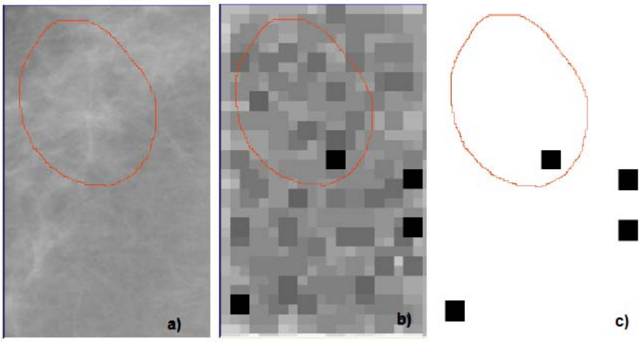


Fig. 12. The ROI of original mammogram (a) and the *probability map* established by option 2 (b) - it is displayed in inverse scale. The thresholded *probability map* with the black blocks which are characterized by the highest probability of ADs appearance (c).

3 Experiments and Results

Our database of mammograms with ADs consisted of 33 images from digital database DDSM (image resolution $43.5 - 50\mu m$) [4]. We tested only ROIs of c.a. 1000×1700 pixels with both benign and malignant abnormal findings. Using option 1 to create the *probability map*, the efficiency of 88% (29 sites of ADs were detected) was achieved at the $FPR = 4.4$ (per ROI), whereas the sensitivity for option 2 was equal 82% at the $FPR = 2.64$. These results are promising for further research. In comparison with the initial efficiency of Gabor filtering stage in ArDist method [7], we achieved sensitivity improvement up to 9%. At the same time the number of false positives per image was significantly reduced.

4 Conclusions

Gabor filtering was checked as the extractor texture orientations. However, ADs are very difficult to detect because of indefinite characteristic features. If we use only Gabor filtering to detect ADs, we should use the *probability maps* established by option 2. In case of the *probability maps* established by option 1 the high number of false positives per image (per ROI in our study) can be reduced in ADs recognition stage on selected ROIs by texture orientation enhancement, analysis and extraction. Moreover, the effective directional features for differentiating ADs from normal breast tissue are searched. Two dimensional transformation with nonseparable, flexible kernels can be applied. Families of countourlets, curvelets, beamlets etc. can be potentially useful for further ADs extraction.

References

1. Ayres, F.J., Rangayyan, R.M.: Characterization of architectural distortion in mammograms. IEEE Engineering in Medicine and Biology Magazine (2005)
2. American College of Radiology (ACR): Breast Imaging Reporting and Data System BI-RADS, 3rd edn. American College of Radiology (1998)
3. Bandodkar, P., Birdwell, R.L., Ikeda, D.M.: Computer-aided detection with screening mammography in a University Hospital Setting. RSNA 236, 451–457 (2005)
4. University of South Florida: Digital Database for Screening Mammography (DDSM). Florida, USA, http://www.marathon.csee.usf.edu/~Mammography/Database.html
5. Dziukowa, J.: Mammografia w diagnostyce raka sutka. Warszawa (1998)
6. Endo, T., Fujita, H., Hara, T., Ichikawa, T., Iwase, T., Matsubara, T.: Automated detection method for architectural distorion areas on mammograms based on morphological processing and surface analysis. Medical Imaging: Image Processing 5370 (2004)
7. Jasionowska, M., Przelaskowski, A., Rutczynska, A., Wroblewska, A.: A two - step method for detection of architectural distortions in mammograms. In: Piętka, E., Kawa, J. (eds.) Information Technologies in Biomedicine. AISC 69, pp. 73–84. Springer, Heidelberg (2010)

8. Ma, W.Y., Manjunath, B.S.: Texture Features for Browning and Retrieval of Image Data. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(8) (1996)
9. The R2 Technology's Image Checker: WWW home page, http://www.r2tech.com
10. Ayres, F.J., Rangayyan, R.M.: Gabor filters and phase portraits for the detection of architectural distortion im mammograms. In: Medical and Biological Engineering, pp. 883–894. Springer, Heidelberg (2006)
11. Ayres, F.J., Leo Desautels, J.E., Prajna, S., Rangayyan, R.M.: Detection of architectural distortion in prior screening mammograms using Gabor filters, phase portraits, fractal dimension and texture analysis. Int. J. CARS (2008)
12. Bovik, A.C., Markey, M.K., Sampat, M.P., Whitman, G.J.: Evidence based detection of spiculated masses and architectural distortions. SPIE, Medical Imaging: Image Processing 5747, 26–37 (2005)

Author Index